Condensed matter physics

Dr. R.J. Cole
Department of Physics and Astronomy, University of Edinburgh

October 9, 2002

Chapter 0

Introduction

0.1 What is Condensed Matter Physics?

Matter is the physical expression of the laws of nature. In this course we aim to explain the properties of 'condensed' matter, by which we mean matter with a density of $\sim 10^{-3}$ kg m^{-3}. Given the enormous diversity of materials, this is a tall order indeed, and so most of the time we will have to be content with a qualitative or semi-quantitative approach for just a few properties. I hope that you will be convinced that condensed matter physics is intellectually demanding, reasonably interesting, and is still an extremely active research area, particularly at Edinburgh.

There are three main components to the course:
- Atomic structure (i.e. the spatial arrangement of atoms),
- Lattice dynamics (i.e. how the atoms move), and
- Electronic structure (i.e. how the electrons behave in condensed matter).

Rather than present a systematic exposition of each, here we adopt an idiosyncratic approach. We will gain insight into mechanical, thermal and optical properties but will focus on the material property that shows the greatest variation, namely electrical conductivity, σ: that of the best conductors is about 10^{30} times that of the worst. If we understand anything about condensed matter then we must be able to explain this.[1]

So profound is the variation in σ that one might suspect that completely different conduction mechanisms are involved in different materials, and it is so. The best electrical conductors, which we call metals, offer the simplest starting point. We may suppose that for an electrical current to flow through a metal there must be a collection of mobile electrons within these materials, and we will pursue this notion in the next chapter. The casual explanation for the existence of insulators is that their electrons are tied up in rigid chemical bonds. We will see that this is a rather superficial (and in fact misleading) explanation.

[1] It is worth noting at the outset that many of the properties of condensed matter appear to be linked. For example those solids which crystallise with what we call the diamond crystal structure tend to be physically hard, electrically insulating with high melting points. Elements which exhibit close-packed structures are usually good conductors of heat and electricity, shiny, ductile, and rather soft.

0.2 Some preliminary considerations

0.2.1 A very many body problem

Whether by classical or quantum mechanics, we are very good at solving problems where a single particle is moving in an external potential (e.g. an electron moving in an electric field) or when two particles interact with each other (e.g. the earth going round the sun). But we cannot in general *solve* the equations of motion for a system of only three interacting particles. Condensed matter comprises lots of particles, all of which are interacting with each other. Although we understand these interactions (and of the fundamental forces we shall have need only of the electromagnetic interaction) this "many-body" nature makes theoretical treatments hard. How then are we to approach condensed matter systems? A good physicist must master the art of carefully applying approximations.

One approach is to focus on the movement of one particular particle. We can then assume that all the other particles give rise to some average potential which acts upon our chosen particle. We can then compute the movement of this particle in the presence of this potential (which we must guess since we don't know how all the other particles move of course). In this way we are approximating the behaviour of the many-body system by that of a collection of single particles which, though interacting, move independently. It turns out that this simplification enables us to understand and explain many things. But occasionally we find phenomena that can only be explained by looking at some of the correlation effects inside the many-body system. Superconductivity is one such example. Who would have guessed that an apparent lack of electrical resistance in a metal could be produced by electrons attracting each other? We all know that electrons repel each other, but in a superconductor they appear to bound together in pairs.

Such effects only emerge from the many-body system. We cannot predict them by looking at how two particles interact or how a single particle moves in an external field. This disturbs the reductionist approach of trying to understand the universe by dissecting it.

0.2.2 Quantum or classical, particle or wave?

Quantum mechanics can be tough, so can we get away without it? We know that when a beam of light is fired at a narrow slit then diffraction effects can be observed - the wave nature of light is manifest. In fact the de Broglie relation tells us that any particle with momentum p can be thought of as a wave with wavelength $\lambda \sim h/p$. But when we walk through a doorway we tend not to notice any diffraction. So how can one decide if a particle will behave classically,[2] or if its wave nature will dominate (requiring us to solve the Schrödinger wave equation)?

The important point is whether the de Broglie wavelength is comparable in magnitude with the relevant physical size for the situation in question (which might be the width of a slit, the separation between two particles etc.). If the wavelength is negligible (e.g. when you walk through a doorway at 1 m s^{-1})

[2] By "classical" we mean "non-quantum mechanical" here. We will probably not encounter any relativistic effects.

then classical physics works, if not, then we must deal with waves. You might be a bit worried that if we sufficiently reduce the momentum of an object as it approaches an aperture, then it would always be possible to make the de Broglie wavelength significant. However it is not always be possible to make the de Broglie wavelength arbitrarily large for a variety of reasons, and here we can list three.
Firstly, there may be no suitable allowed states.
Secondly, an object with temperature T will have a random motion with energy of order $\sim k_b T$, where $k_b = 1.38 \times 10^{-23}$ J K^{-1} and is known as Boltzmann's constant, and hence a "thermal wavelength" of $\sim h/\sqrt{2mk_bT}$. The thermal wavelength of a snooker ball at room temperature is $\sim 10^{-23}$ m, which is extremely small on the length scale of the ball, and we confirm that wave effects (i.e. quantum mechanics) should not be manifest during a game of snooker, no matter how slowly you strike the balls. On the atomic scale, where here the natural length is 1 Å $= 10^{-10}$ m, the same argument implies that at room temperature the molecules in water are bumbling around like classical particles, while the electrons within the molecules demand a quantum mechanical treatment. It may seem that reducing T will eventually cause the de Broglie wavelength to explode, but there is a third point.
The Uncertainty Principle requires that the de Broglie wavelength cannot exceed Δx, the spatial uncertainty. This implies that atoms in condensed matter can usually be treated classically.

0.2.3 Statistics

Quantum mechanics is not simply a matter of finding the eigenstates of a Hamiltonian. When we consider a system of many particles we must think about statistics. Take, for example, the electrons in a piece of metal. If, using some approximation, we describe this many body system using a collection of single electron states, it still remains for us to say how many electrons will go into which states. At $T = 0$ they will try to drop as low in energy as they can, and we call this the ground state. For $T > 0$ there will inevitably be thermal excitations, and we also need to understand how this affects the occupations. In classical systems the probability that a particular state will be occupied is given by Maxwell-Boltzmann statistics but in quantum mechanical systems things are more complicated. It is a remarkable fact that all systems of particles with half-integer spin obey one kind of statistics (Fermi-Dirac statistics), while those consisting of integer spin obey another (Bose-Einstein statistics). This "spin-statistics theorem" is universal - no exceptions are known.

Already we have touched upon some rather profound issues and it should be clear that if we are to get to grips with condensed matter then we will need to keep our wits about us. Since condensed matter constitutes the raw materials for all manner of industries from food production to microelectronics to ship building, the technological imperative for mastering this field can scarcely be overstated.

Chapter 1

The free electron model of metals

It is perhaps surprising that a hundred years ago there was virtually no understanding of the physics of solids. The origins of modern condensed matter physics can be traced back to attempts to account for the properties of metals. We will briefly review the classical and semi-classical free electron theories of the metallic state dating from the turn of the century, and we will see both successes and failures. The free electron model provides a meaningful baseline against which the properties of real metals can be judged. More importantly, *understanding* the successes and failures of the free electron model will highlight the central issues of this course.

1.1 The classical free electron gas: The Drude-Lorentz model

At the turn of the century, Einstein had not yet explained the photoelectric effect, Rutherford had not determined the size of the nucleus, Bohr had not speculated on the discrete nature of electronic "shells" in atoms, and the formulation of quantum mechanics was still decades away. Although the structure of the atom was not known, Thomson had discovered the electron (1897), enabling Drude (1900) and Lorentz (1905) to formulate a model to explain two of the most striking properties of the metallic state, namely the conduction of electricity and heat.

1.1.1 The physical model: collisions

Drude attributed conduction in the metallic state to the most loosely bound electrons in atoms somehow becoming mobile. In fact these "conduction electrons" were assumed to move freely through space apart from collisions, not with each other but rather with the much larger atomic cores, as shown in Fig. 1.1 on the following page. Geometric considerations imply a mean free path between

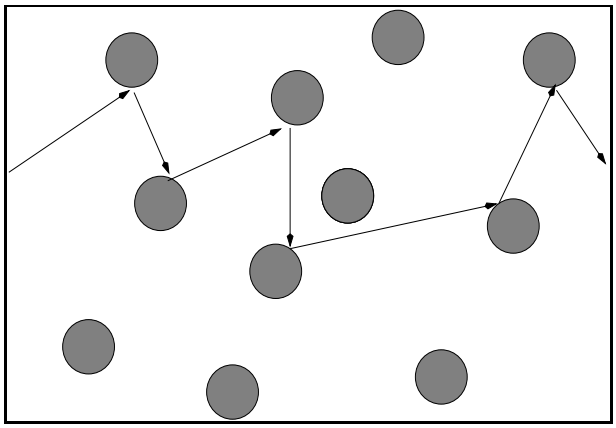

Figure 1.1: The physical interpretation of the Drude model.

collisions of

$$\ell \sim \frac{1}{\pi R^2 N}, \tag{1.1}$$

where R is the core radius and N is the number of atoms per unit volume. Guessing that R is not that much smaller than the atomic radius, Drude deduced ℓ to be a few Å. He assumed that after a collision a conduction electron has random direction and a speed which **does not depend on its velocity beforehand**, but is determined by the *equilibrium distribution function.*

At this time, the most natural treatment of free electrons was provided by the kinetic theory of the ideal gas, and so by analogy the collisions were assumed to be instantaneous and physical (billiard ball like). Since each atom contributes of the order of one conduction electron, the electronic density n is of the same order of magnitude as N ($\sim 10^{29}$ m^{-3}). Notice that the free electron gas is a "fake" many body problem. Although they are many in number these are non-interacting electrons, each moving independently and occassionally suffering a collision with an ion core. Within these assumptions we are entitled to treat them one at a time.

• So what happens to the electron-electron and electron-nucleus interactions?

We can postpone answering this question for a little while since the first job of a model is to correctly reproduce observed behaviour. Only then do we have to explain *why* the model works.

A classical free particle is allowed to have any velocity. By analogy with the ideal gas, the number of conduction electrons per unit volume with velocities in the range $\underline{v}$ to $\underline{v} + d\underline{v}$ is $n^{\underline{v}}_{\rm MB} d\underline{v}$ in the Drude-Lorentz model, where $n^{\underline{v}}_{\rm MB}$ is the Maxwell-Boltzmann (MB) equilibrium velocity distribution function:

$$n^{\underline{v}}_{\rm MB}(T) = n \left(\frac{m}{2\pi k_b T} \right)^{3/2} \exp\left(\frac{-mv^2}{2k_b T} \right) \tag{1.2}$$

where m is the electron mass, and n is the total number of conduction electrons per unit volume, irrespective of their energies. Note that this expression has the

form of a normalising constant multiplied by a Boltzmann factor. In "velocity space" (or "k-space", where $\underline{k} = m\underline{v}/\hbar$), the MB distribution is spherically symmetric and centred on the origin.

If we are only interested in the average *speed* $v = |\underline{v}|$, then the distribution function becomes

$$n^v_{\rm MB}(T) = n\left(\frac{m}{2\pi k_b T}\right)^{3/2} \exp\left(\frac{-mv^2}{2k_b T}\right) 4\pi v^2. \tag{1.3}$$

Using this expression Drude calculated the average electronic speed is $\sim 10^5$ m/s at room temperature and deduced that τ, the average time between collisions (also called the *relaxation time*) is between 10^{-14} and 10^{-15} s.

If the electron gas is subject to some external force $\underline{F}$, then the electrons accelerate. However, in the Drude-Lorentz model an electron emerging from a collision does not "remember" if it had previously been accelerated or not. For a constant applied force the repeated "resetting" of the electron velocities prevents the electrons accelerating indefinitely but results in the establishment of a drift velocity $\underline{v}_d$ in the direction of the applied field, superimposed on the random thermal motion of the conduction electrons. (In k-space this amounts to shifting the spherically symmetric distribution away from the origin.) Since the collisions are effectively a means of damping, the equation of motion for the drift velocity is

$$m\frac{d\underline{v}_d}{dt} = \underline{F} - \frac{m\underline{v}_d}{\tau}. \tag{1.4}$$

1.1.2 Electrical conductivity

Consider the electron gas in a spatially invariant DC electric field $\underline{E}$. When a steady state is established, $d\underline{v}_d/dt = 0$ and the electronic drift velocity is constant, corresponding to the current density $\underline{J} = -ne\underline{v}_d$. It follows that the current density is proportional to the applied field $\underline{J} = \sigma_0\underline{E}$ (i.e. Ohm's law), where the DC conductivity σ_0 is given by

$$\sigma_0 = \frac{ne^2\tau}{m}. \tag{1.5}$$

Using their estimate of the relaxation time τ, Drude and Lorentz obtained conductivities which were in quite good agreement with experiment. Notice that without collisions, $\tau \to \infty$ and so the conductivity becomes infinite.

In the presence of an AC field $\underline{E}(t) = \underline{E}_0 e^{-i\omega t}$, the steady state drift velocity must oscillate at the same frequency as the applied field, although not necessarily in phase with it. It is easily shown that the AC conductivity is then

$$\sigma_\omega = \frac{\sigma_0}{1 - i\omega\tau}. \tag{1.6}$$

Playing around with some basic equations from optics and electromagnetic theory we will see in a Problem Sheet that free electron metals should be highly reflective in the visible region of the electromagnetic spectrum, but transparent to ultra-violet. This is indeed the case for many metals.

1.1.3 Thermal conductivity

The electrical conductivities of metals and insulators are profoundly different. When it comes to thermal conductivity the distinction is much less dramatic, but metals do tend to conduct heat about one hundred times better than insulators at room temperature. This is an early indication that there is a mechanism for thermal conduction which is not related to electrical conduction. But it also suggests that when metals conduct heat the electrons are doing most of the work.

The electron gas model of metals explains thermal conduction as follows. Imagine holding a metal bar in a fire. The electrons in the end that is in the fire tend to be travelling much faster than those in the end which we are holding. But some of the fast ones may happen to be travelling towards the end we are holding. They travel a certain distance before being scattered in some random direction. Thus they carry thermal energy along the bar and before long it becomes too hot to hold. In the Drude-Lorentz model the heat flow is proportional to the tempertaure gradient[1] and the constant of proportionality, i.e. the thermal conductivity κ of the free electron gas, turns out to be

$$\kappa = \frac{2}{3}\frac{\bar{\epsilon}\tau C_V}{m} \tag{1.7}$$

where $\bar{\epsilon}$ is the average electron energy and C_V is the heat capacity of the electron gas (the rate at which the energy density of the solid changes with T, keeping the volume constant).

To calculate $\bar{\epsilon}$ we change the variable in Eq. 1.3 from speed to energy, obtaining the equilibrium energy distribution function

$$n^{\epsilon}_{\mathrm{MB}} = \frac{2n}{\sqrt{\pi}\,(k_b T)^{3/2}}\sqrt{\epsilon}\exp(-\epsilon/k_b T) \tag{1.8}$$

also known as the *density of occupied energy levels per unit volume*, or simply the *density of occupied energy levels*. The average electron energy is then

$$\bar{\epsilon} = \frac{1}{n}\int_0^\infty \epsilon\, n^{\epsilon}_{\mathrm{MB}}(T)d\epsilon = \frac{3k_b T}{2} \tag{1.9}$$

(the classical "equipartition" result for three degrees of freedom). Since the electrons do not interact, the total energy of the electron gas is equal to the sum of energies of the individual electrons and it follows that

$$C_V = n\left(\frac{\partial\bar{\epsilon}}{\partial T}\right)_V = \frac{3nk_b}{2}, \tag{1.10}$$

and so

$$\kappa = \frac{3k_b^2 nT\tau}{2m}. \tag{1.11}$$

Eq. 1.11 gave thermal conductivities which again were in quite good agreement with experiment.

[1] This is known as "Fourier's law" and is the thermal equivalent of Ohm's law.

Better still, Eqs. 1.5 and 1.11 gave a "Lorenz number" L, defined by $\kappa/(\sigma T)$, of

$$L = \frac{3}{2}\left(\frac{k_b}{e}\right)^2 . \tag{1.12}$$

This explained the empirical observation of Wiedemann and Franz (1853) that L is independent of T and varies very little from metal to metal. The equation also reproduces the observed numerical value of L.

1.1.4 Failure of the model

Hidden behind this great triumph of the Drude-Lorentz model there was a difficulty: it predicts the contribution to C_V due to free electrons is $3nk_b/2$. In fact C_V has this approximate magnitude for both metals and insulators.[2] It must be concluded that (i) there is a contribution to the heat capacities of all solids which we haven't yet identified, and (ii) the electronic contribution predicted by the Drude-Lorentz model for metals is not observed. The problem was a serious and stubborn one and cast a shadow over the free electron model for the next twenty years.

In fact there was an older puzzle. The work of Faraday, Ampère, Lenz etc. had shown that a current-carrying wire experiences a force (the *Lorenz force*) when placed in a magnetic field. In 1879 Hall tried to determine whether this force acted on the wire as a whole or on some substituent responsible for the electric current. He predicted that the force would act upon the substituent (i.e. the electrons) and, since their path length through the wire would be parabolic instead of straight (and therefore longer), an increase in electrical resistance should be observed. Hall failed to observe what we now call *magneto-resistance* (which is usually very weak), but found that a voltage built up across the wire perpendicular to the direction of the current.

Hall quickly realised he had overlooked something. Initially the electrons are drawn sideways by the magnetic field but then they build up on the edge of the wire. This creates an electrostatic repulsion that eventually balances the Lorenz force, preventing the transverse motion of further electrons. Having explained the origin of this transverse voltage, Hall was in for another surprise. The 'Hall voltage' observed for Au, Cu, K and Na suggested that these metals have approximately one free electron per atom. This all fitted with clues from chemistry about the "monovalent" nature of these elements. But the Hall voltages for the divalent metals Mg, Cd and Be and the trivalent metal Al were found to have the wrong sign, as if the current were being transported by positively charged particles. This is not in keeping with our picture of conduction by free electrons.

[2] It is generally found that $C_V \sim 3Nk_b$ works well for both metals and insulators.

1.2 Sommerfeld model of the free electron gas

We observed in the first lecture that there are only three basic ingredients to the free electron model of the metallic state: the electronic states, the statistics of how those states are occupied, and the damping effect caused by the conduction electrons being scattered. The Drude-Lorentz model (based upon classical physics) had some success but could not be reconciled with all the experimental facts. Where did we go wrong? In this section we will start again but this time using a quantum mechanical approach. We'll do more of the maths but don't be put off by that - there is only one important new idea. (The *Pauli principle*.)

When puzzles like the black-body problem of electromagnetism and the heat capacity problems for solids arose in the nineteenth century, doubts were raised about the understanding of statistical mechanics and thermodynamics. It was clear that the "equipartition of energy" derived from the Boltzmann law[3] was not universal. We now believe that the Boltzmann law *does* hold, the problem lay with the nature of matter itself.

The 1920's saw the advent of a quantum mechanical description of atoms. The de Broglie wavelength for an electron with energy 13.6 eV (the kinetic energy of an electron in a hydrogen atom) is a few Å and this is of the same magnitude as the interatomic separations in condensed matter. On this basis one might expect that electrons in solids will demonstrate diffraction and interference. Perhaps we need to take this seriously. [4] Maybe we should also consider the many body nature of the electron gas. In the "one-electron approximation" one replaces the true (unknown) wavefunction of the whole system which we denote $\Psi_N(\underline{r}_1, \underline{r}_2, \cdots, \underline{r}_N)$ by an approximate one built up from states which each describe only one electron: $\psi_1(\underline{r}_1), \psi_2(\underline{r}_2), \cdots, \psi_N(\underline{r}_N)$). Maybe we should be aiming to get Ψ_N rather than considering each electron separately.

We will return to these two issues when we get a bit more sophisticated, but Sommerfeld brushed them aside by continuing to assume that the conduction electrons in a metal do not interact with each other or with the ions. If they don't feel the ions, they can't be diffracted by them. If they can't feel each other, then the one-electron approximation is exact. Sommerfeld pointed out that there is something more fundamental to grasp: the *quantum statistics theorem*. This states that the wavefunction of a system of identical spin-half particles must be anti-symmetric with respect to interchange of any two of the substituent particles. Within the one-electron approximation it follows that no two electrons can be in the same quantum state, and we usually call this the *Pauli principle*. In a rather mysterious way the electrons in a metal would be somehow aware of each other even if there were no Coulomb force between them.

We now need to (i) figure out what are the states (using quantum mechanics), then (ii) make sure we put only one electron in each.

[3] This states that the probability of a system being in a state which has energy E is proportional to $e^{-E/k_b T}$.

[4] In a few weeks we should realise that this is possibly the most important observation in the whole course - it will explain the occurrence in nature of metals, semiconductors and insulators.

1.2.1 The eigenstates

Starting with the time-dependent Schrödinger equation we write the wavefunction of a free electron as the product of a spatial function and a temporal function: $\psi(\underline{r})T(t)$. It is straight forward to show that $T(t) = \exp(-i\epsilon t/\hbar)$ where ϵ and ψ are the eigenenergy and wavefunction in the time-independent Schrödinger equation. States of this form are termed *stationary states* since their probability density $(\psi T)(\psi T)^*$ equals $\psi\psi^*$ and is independent of time.[5] For a free particle the time-idependent Schrödinger equation is simply

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\underline{r}) = \epsilon\psi(\underline{r}) \tag{1.13}$$

and one can easily show that

$$\psi_{\underline{k}}(\underline{r}) \propto e^{i\underline{k}.\underline{r}} \tag{1.14}$$

are eigenfunctions with energy $\epsilon(\underline{k}) = \hbar^2k^2/2m$. Notice that the full wavefunction of the free electron has the form $\exp i(\underline{k}.\underline{r} - \epsilon t/\hbar)$, the equation for a plane wave with wavevector[6] $\underline{k}$. There are infinitely many such solutions of the Schrödinger equation and so we have distinguished them by writing $\underline{k}$ as a subscript on the wavefunction symbol. $\psi_{\underline{k}}(\underline{r})$ is also an eigenfunction of the momentum operator: $-ih\nabla\psi_{\underline{k}}(\underline{r}) = \hbar\underline{k}\psi_{\underline{k}}(\underline{r})$, and so each electron has well-defined momentum given by $\hbar\underline{k}$.[7] Just as in the classical treatment, we have $\epsilon = p^2/2m$ and any momentum is allowed.

All pieces of metal are finite in size so one might wonder what happens to the electrons as they encounter a surface. In any reasonably sized sample surface effects are negligible and we can largely avoid them by imposing periodic boundary conditions. Consider a macroscopic piece of metal in the shape of a cube with linear dimension L and volume $W = L^3$. Periodic boundary conditions mean that if an electron passes through any particular face of the cube it immediately re-enters it cube through the opposite face. Mathematically this means

$$\psi_{\underline{k}}(x + L, y, z) = \psi_{\underline{k}}(x, y, z) \qquad \text{etc.} \tag{1.15}$$

Combining Eq. 1.14 and 1.15 we find that the components of $\underline{k}$ must be of the form

$$k_x = \frac{2\pi}{L} \times \text{integer} \qquad \text{etc.} \tag{1.16}$$

Imposing boundary conditions has introduced an artificial restriction on the allowed $\underline{k}$ vectors (and hence momenta): they are evenly distributed with a spacing of $2\pi/L$ in the x, y and z directions. But by choosing L to be very large we can ensure that the allowed $\underline{k}$ vectors are arbitrarily close together.

[5] Stationary states have a precise energy and an infinite lifetime in accord with the energy-time form of the uncertainty principle $\Delta\epsilon\Delta t \gtrsim \hbar$.

[6] The wavevector of a plane wave has magnitude $2\pi/\lambda$, where λ is the wavelength, and points in the direction of propagation of the wave.

[7] We don't need to make a fuss about it but there is an important physical principle here: for every symmetry there is a corresponding conservation law. Whenever a particle moves in a constant potential (i.e. the potential is translationally invariant) then its linear momentum is conserved. Similarly, if a particle moves in a central potential (i.e. the potential is spherically symmetric) then its angular momentum is conserved.

A good thing about having a discrete set of allowed states is that it is easier to keep track when we have to count them. The volume (of k-space) around each allowed $\underline{k}$ point is $(2\pi)^3/W$, and the density of allowed momentum states is therefore

$$g^{\underline{k}} = \frac{W}{(2\pi)^3}. \tag{1.17}$$

Clearly it is uniform, much the same as for the classical electron gas. The corresponding density of energy levels is

$$g^{\epsilon} = \frac{W}{4\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}\sqrt{\epsilon}. \tag{1.18}$$

It is conventional at this stage to multiply this result by a factor of 2 to account for spin degeneracy of the states, and to divide by W to give the density of states *per unit crystal volume*.

1.2.2 Ground state of the quantum free electron gas

Now let's consider how the states are occupied, starting with the $T \to 0$ limit which we call the ground state. In the classical electron gas the velocities (and ϵ and k) collapse to zero, but the Pauli principle allows only two electrons (with opposite spin) to have $\underline{k} = 0$ in the quantum electron gas. We must place the rest of them in successively higher energy states. The highest occupied level we call the *Fermi level*, denoting its energy ϵ_f and the magnitude of its wavevector k_f. In k-space the levels are now uniformly occupied for $k \leq k_f$ and unoccupied outside this *Fermi sphere*. The Fermi energy is determined simply by the number of conduction electrons per unit volume and the available density of states. For most metals ϵ_f is of the order of a few eV (where 1 eV $= 1.6 \times 10^{-19}$ J) and $k_f \sim 10^{11}$ m$^{-1} = 10$ Å^{-1}, corresponding to a Fermi velocity of $v_f \sim 10^6$ ms^{-1}. It is vital to realise v_f is the typical electron velocity in the quantum free electron gas at $T = 0$. It should be clear that the properties of the free electron gas are dominated by the Pauli principle and in *this* sense the free electron gas is fundamentally quantum mechanical.

1.2.3 The heat capacity

When $T > 0$ there is a finite chance that the available thermal energy will excite an electron into a state above the Fermi level. You will have seen last year that the probability of a state with energy ϵ being occupied is now given by $f^{\epsilon}_{\mathrm{FD}}$, the Fermi-Dirac factor:

$$f^{\epsilon}_{\mathrm{FD}}(T) = \frac{1}{\exp\left[(\epsilon - \mu(T))/k_{\mathrm{B}}T\right] + 1}. \tag{1.19}$$

For our purposes we can take the *chemical potential* $\mu(T)$ to be equal to the Fermi energy ϵ_f.[8] It is easy to show that $f^{\epsilon}_{\mathrm{FD}}(T) \sim 1$ for $\epsilon \lesssim \epsilon_f - k_b T$, and $f^{\epsilon}_{\mathrm{FD}}(T) \sim 0$ for $\epsilon \gtrsim \epsilon_f + k_b T$.

[8] Strictly speaking it is determined by the requirement $\int_0^\infty n^{\epsilon}_{\mathrm{FD}}(T)d\epsilon = n$. In fact, $\mu(0) = \epsilon_f$, and $\mu(T)$ varies negligibly for temperatures below the melting points of most metals.

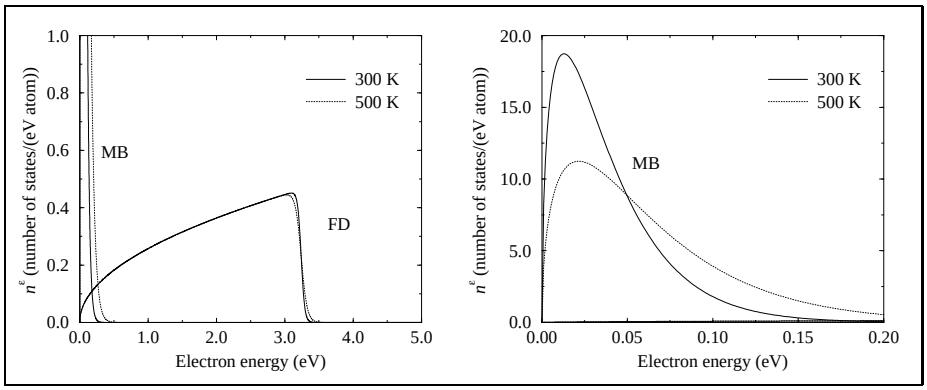

Figure 1.2: Comparison of the density of occupied states at room temperature (solid curves) and $T = 500$ K (dotted curves) for an electron gas with the density of metallic sodium. MB and FD refer to Maxwell-Boltzmann and Fermi-Dirac statistics respectively.

At finite temperature the density of *occupied* states in the quantum electron gas is

$$n^\epsilon_{\rm FD}(T) = g^\epsilon f^\epsilon_{\rm FD}(T) \tag{1.20}$$

where g^ϵ is just the density of states given by Eq. 1.18. At everyday temperatures (say less than 10^4 K) the Fermi-Dirac distribution is spectacularly different from the Maxwell-Boltzmann distribution.[9] $n^\epsilon_{\rm FD}$ and $n^\epsilon_{\rm MB}$ are compared in Fig. 1.2.

Recalculating $\bar{\epsilon}$ (Eq. 1.9 on page 8) with FD statistics, Sommerfeld found that the heat capacity of the quantum free electron gas to be

$$C_V = \frac{\pi^2}{2}\left(\frac{k_{\rm B}T}{\epsilon_f}\right) n k_{\rm B}. \tag{1.21}$$

Since the Fermi energy is so much larger than usual thermal energies (at room temperature $k_{\rm B}T \sim 0.02$ eV), it is evident that the Sommerfeld model predicts a much lower heat capacity than its classical counterpart (typically a factor of ~ 50 smaller at room temperature).

• What is the physical origin of this result?

Increasing the temperature of the classical electron gas moves the whole distribution of occupied states to higher energy, as can be seen in Fig. 1.2. In the quantum gas only those very few electrons within $\sim k_{\rm B}T$ of the Fermi energy can be thermally excited because of the Pauli principle. At low temperature the Sommerfeld model gives a reasonable estimate of the heat capacity of real metals, although the experimental data for potassium in Fig. 1.3 on the following page reveals a contribution to C_V which appears to scale with T^3. Near room temperature the heat capacity of most solids is of order nk_b and the electronic

[9] When $k_bT \gg \epsilon_f$, $n^\epsilon_{\rm FD}$ converges towards $n^\epsilon_{\rm MB}$, i.e. we approach the classical regime. We might have deduced this by considering the electron thermal wavelength for $T \gg \epsilon_f/k_b$.

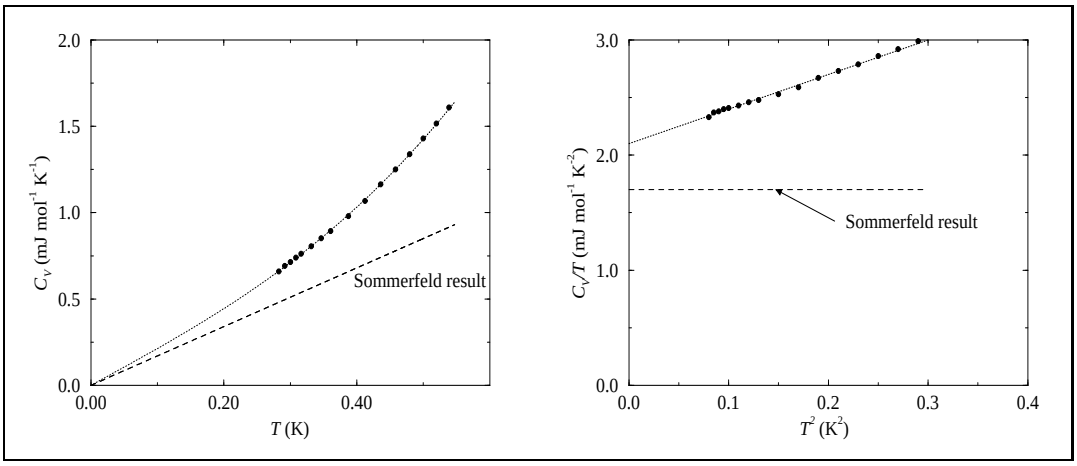

Figure 1.3: Temperature dependence of the heat capacity C_V of Potassium. The linear relation of C_V to T predicted by the Sommerfeld model is quite well reproduced by experiment at low temperature. However a cubic contribution (which dominates at higher temperature) can be seen by plotting C_V/T against T.

contribution is completely swamped. Clearly there is another type of effect here, but at least we now understand the electronic contribution to the heat capacity of metals.

1.2.4 Electrical conductivity

It is worth emphasizing that although Sommerfeld solved the Schrödinger equation to calculate the stationary states of the free electron gas, he appealed to a collision mechanism to explain electrical resistance just as his predecessors had done. Since the Drude estimate of the collision time seemed to work (it gave the correct room temperature electrical conductivity), this value $\tau \sim 10^{-14}$ s was retained by Sommerfeld. But in the Sommerfeld model the average electron speed at this temperature is about 10 times greater than Drude and Lorentz had calculated. We must therefore conclude that λ, the average distance travelled between collisions, is ~ 10 times longer than Drude and Lorentz had supposed and this must cast doubt over the collision mechanism they had advanced.

We should abandon the Drude collision mechanism at this stage, but *something* must scatter electrons, and experimental measurements provide a few small clues. It was well known that the conductivity of metals decreased steadily with T. It was also found that it decreased suddenly upon melting (at constant T). The conductivity of pure metals was found to be reduced when impurities were added. We're not quite ready to digest this information yet, but we'll try to come back to it.

Sommerfeld's retention of a classical description of electron dynamics (Eq. 1.4) requires some justification. How do the wave-like eigenstates of the quantum mechanical free electron gas relate to the particle picture implicit in classical dynamics? We'll tackle this matter in "Solid state physics" (SSP) next term, but in case you are interested: we must identify each electron with a *wavepacket*

of free electron waves. These are spatially localized on the scale of the collision length, but are delocalized on the scale of the atom.

1.2.5 Thermal conductivity

How did Drude and Lorentz get good values for κ when their calculated value of the electronic contribution to C_V was so poor? By looking at Eq. 1.7 on page 8 we see that Drude and Lorentz were very fortunate. By using MB statistics they overestimated C_V but also underestimated the average electron energy by a similar factor. When Sommerfeld corrected C_V and $\bar{\epsilon}$ he obtained more or less the same κ as Drude and Lorentz, which had already been found to be in agreement with experiment.

1.2.6 Summary

There are three main pieces to the Sommerfeld theory of the electron gas:

- Solution of the Schrödinger equation to get the allowed electronic states.

- The use of Fermi-Dirac statistics to determine the population of the states.

- Electron scattering: Appeal to some (unknown) scattering mechanism, and classical electron dynamics (between collisions).

After the work of Sommerfeld the free electron model was in much better shape. The electical conductivity, thermal conductivity, heat capacity and reflectivity were all quite well described, but a number of questions remained:[10]

What is(are) the true mechanism(s) for electron scattering?
How can the conduction electrons in a metal move seemingly unhindered over such long distances? (λ in pure metal crystals can be increased to centimetres by lowering T.)
What is the dominant contribution to the heat capacities of solids for $T \gtrsim 10$ K?
Why does conduction of electricity in some metals appear to take place via the transport of positively charged particles?

To find the answers and to arrive at a more satisfactory description of metals (and condensed matter in general) we must now consider what happens if the electrons are allowed to interact with the ion cores and with each other. To do this we need to start taking an interest in where the ions in a metal are. We turn therefore to the subject of atomic structure.

[10] To the questions listed below we might add a number of deeper questions: Why are some elements metals and some not? What holds a metal together if the valence electrons do not form localized bonds like in a molecule? Why are some electrons free in the first place while others remain bound to the ion cores? These are difficult questions to answer properly, but we'll try in SSP.

Chapter 2

Structure at the Atomic Scale

The discovery of electrolysis by Faraday in the nineteenth century was quite momentous and showed that "chemistry is electricity". When atoms come together and form chemical bonds it is electrons that are responsible. Given the pre-eminent position the electron seems to have, our ultimate aim is to construct a respectable theory of the electronic structure of condensed matter. Having introduced the free electron model we now have to account for the interactions between electrons and the electron-ion interactions. This is difficult, so we will start to focus increasingly on the simplest form of condensed system, namely the perfect crystal.[1] But obviously we need to take an interest in atomic scale structure, i.e. where the atoms actually are.

But before we focus down in the next chapter, let's think more generally about the interactions present in condensed systems and (qualitatively at least) the relationship between chemical bonding and structural form. As well as considering ground state structures we will briefly consider *phase diagrams* and the nature of phase transitions. In §2.3-2.5 we will develop a systematic way of describing the structure of condensed systems, before seeing the primary means of experimentally determining atomic scale structure (x-ray scattering) in §2.6-2.7.

2.1 Chemical Bonding

When we bring various types of atoms together (e.g. a bunch of silicon atoms, a bunch of sodium atoms, a bunch of chlorine atoms, a mixture of sodium and chlorine etc.) rather substantial electronic rearrangements may take place. We usually refer to this as *chemical bonding*, but in all cases it is only the outermost electrons of an atom that can participate.[2] The number and type of bonds that

[1] It turns out that even this is difficult.

[2] The inner electrons remain tightly bound to the nuclei and are often called *core* electrons. On practical grounds we have no choice, but to say that core electrons play *no* role in bonding would not be wholly true. For example, the group 4 elements carbon, silicon, germanium, tin and lead all have the same outer electronic configuration but many different properties. Carbon is very special, having a branch of chemistry all to itself. Subtle differences in the

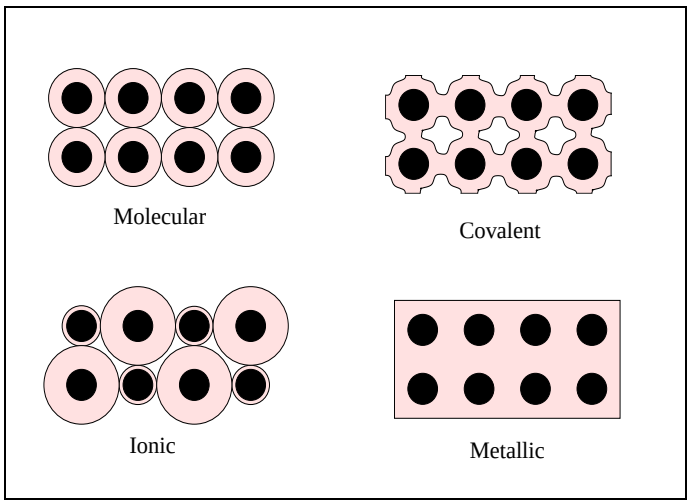

Figure 2.1: Schematic representation of the charge density for the four main types of chemical bonding.

may be formed depends on how many electrons each atom possesses, known in chemical circles as the theory of *valency*. The concept of chemical bonding deals largely with the spatial redistribution of electron density when atoms combine, and this is what we will briefly describe here. Only in the simplest cases can we write down meaningful expressions for interatomic potentials. In fact in some cases we will not even be able to offer much of an explanation why the bonds hold together at all. Some things that we are very familiar with are actually quite hard to explain.

2.1.1 Molecular bonding

We start with the elements on the far right of the periodic table, the "noble gases": helium, neon, argon, krypton, xenon and radon. The atoms of this group of elements are *extremely* inert. We will assume that the electron orbitals that are present when these atoms are isolated remain more or less intact under all circumstances. Since their electronic structure does not change, we can then treat each atom as a composite particle.

An "ideal" gas has an equation of state of the form

$$PV \propto T \tag{2.1}$$

where P is pressure and V volume. You will remember that this doesn't quite work for real gases. Firstly, we must account for the fact that atoms have a finite size and they can't interpenetrate each other. We can try to account for this by making a correction to the V in Eq. 2.1. We also have to fiddle with P since the atoms of a real gas attract each other a little bit. We will squeeze quite a lot out of these two observations.

Consider two noble gas atoms denoted 1 and 2 separated by a distance r. Although each atom is spherically symmetric *on average*, at any particular

cores of these elements are to blame.

instant one of the atoms may have a dipole moment $\underline{p}_1$. This dipole will give rise to an electric field p_1/r^3 at the other atom, and this field will induce a dipole moment with magnitude $p_2 = \alpha p_1/r^3$ in the second atom, where α is the atomic polarisability. The electrostatic energy of the two dipoles is proportional to

$$-\frac{p_1 p_2}{r^3} = -\frac{\alpha {p_1}^2}{r^6}. \tag{2.2}$$

Although the time-average of each dipole is zero, their average interaction energy is non-zero since it depends on ${p_1}^2$ which is always positive. This attraction between fluctuating dipoles is known as the *van der Waals* or *London* force.

Pauli repulsion of the electrons from adjacent molecules prevents molecular solids from collapsing in on themselves. This "force" is extremely strong for small separations and so must have shorter range than the fluctuating dipolar attraction. It is conventional to represent the Pauli repulsion by a potential which varies inversely with the 12th power of r. The total potential energy for two molecules is then given by

$$V(r) = 4V_0 \left[\left(\frac{\ell}{r}\right)^{12} - \left(\frac{\ell}{r}\right)^{6} \right], \tag{2.3}$$

known as the *Lennard-Jones potential*, plotted in Fig. 2.2. V_0 gives the depth of the energy well, while ℓ is a length determining the range of the potential (the minimum of the curve occurs at 1.09ℓ). Detailed analysis of the departure of the noble gasses from ideal behaviour yields the values $V_0 \sim 0.01$ eV and $\ell \sim 3$ Å.

The discussion above allows us to guess the preferred ground state (i.e. $T = 0$) structures of the noble solids.[3] We have seen that the origin of interatomic attraction is the fluctuating dipole-dipole interaction, and this suggests that the noble elements will crystallise with each atom having the maximum possible number of neighbours. The more dipoles around, the greater the attraction and the lower the energy. This is indeed the case - they adopt the face centred cubic crystal structure (which we will discuss soon) where the packing is as dense as possible with each atom having 12 nearest neighbours.

We can also estimate the energy of the ground state of the noble solids by assuming it can be expressed as a sum of interaction energies of Lennard-Jones form for all pairs of atoms in a crystal. We will do this properly in SSP next term, but a good estimate can be made by further assuming that only nearest neighbour pairs make a significant contribution. We deduce the ground state energy of the noble solids to be of order -0.1 eV per atom, where zero corresponds to infinite separation.[4] Experimental measurements verify such estimates, but the calculations are most successful for the heavier noble solids and tend to overestimate the cohesion of the lighter ones.

This discussion has focussed (for simplicity's sake) on the noble gases, but in fact these are a special case of "molecular bonding" in which the molecular subunits are single atoms. There are many other systems (e.g. carbon dioxide, oxygen, methane etc.) which comprise molecules which attract via the van

[3] We are making the reasonable assumption that the ground state will be a solid.

[4] Minus the ground state energy is sometimes called the *cohesive energy*, given these definitions.

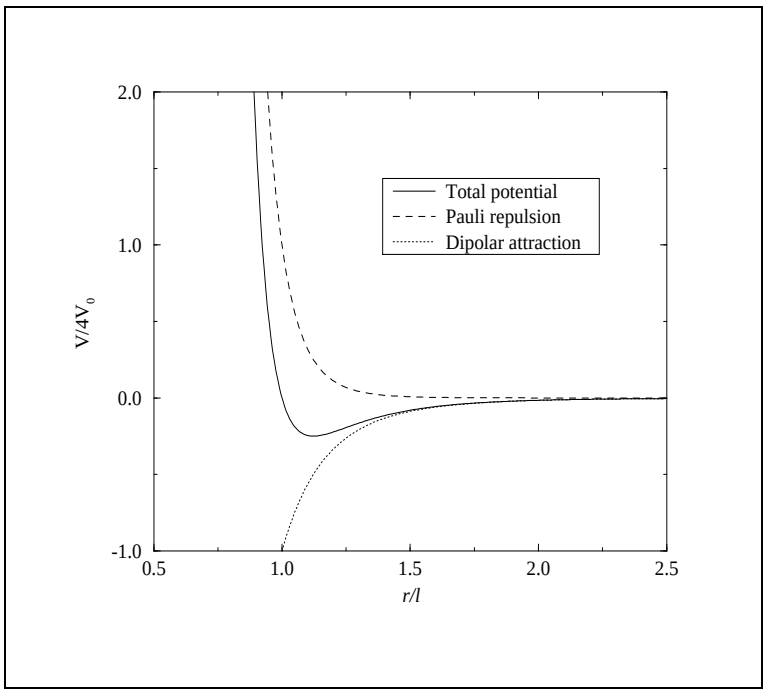

Figure 2.2: The Lennard-Jones potential energy as a function of the separation of two atoms.

der Waals interaction, but whose internal molecular structure[5] is perturbed only negligibly by the weak intermolecular bonds. In general the structure of a molecular solid is determined by the shape of the constituent molecules as well as the drive for close-packing caused by intermolecular van-der-Waals interactions.

2.1.2 Covalent bonding

The distinguishing property of covalent bonding is the enhanced electron density along the lines joining adjacent atoms, usually referred to as *directional bonds*. In the crudest terms it is said that a covalent bond is formed by the sharing of a valence electron with a neighbouring atom which chips in an electron of its own. In covalent systems the charge density of the participating atoms is usually considerably redistributed. Experiment shows that covalent bonds are extremely strong. It is not possible to put together a crude theory of the cohesive energy of covalent bonds, for that we need a quantum mechanical effort,[6] but it is possible to understand the basic idea.

The best starting point for discussing covalent bonding is the electronic structure of molecules. We'll take the simplest possible one: the H_2^+ molecule. This consists of two protons and a single electron. We assume that the protons are fixed at some particular distance apart and our task is to calculate the

[5] The bonds *within* the molecules are strong and usually covalent.

[6] We may return to this later in CMP, but you should see this in "Atomic and molecular physics" or in SSP.

wavefunction of the electron.[7] With these assumptions and some fancy geometry the electron wavefunction of H_2^+ can be solved analytically. We won't bother doing that because (i) it is hard, and (ii) the procedure is not general. If we have more electrons, we are back to insoluble many-body problems so we may as well do an approximate treatment from the start.

The "molecular orbital" (MO) of the electron in H_2^+, which we denote $|\psi_{MO}\rangle$, satisfies the Schrödinger equation:

$$H|\psi_{MO}\rangle = -\frac{\hbar^2}{2m}\nabla^2|\psi_{MO}\rangle + U_{AB}|\psi_{MO}\rangle = \epsilon_{MO}|\psi_{MO}\rangle. \tag{2.4}$$

Let's assume that $|\psi_{MO}\rangle$ can be well approximated by a linear combination of ψ_A and ψ_B, 1s hydrogen wavefunctions centred on proton A and proton B respectively:

$$\psi_{MO} = c_A\psi_A + c_B\psi_B. \tag{2.5}$$

Since the right hand side is a linear combination of atomic orbitals this approach is called LCAO-MO. Inserting Eq. 2.5 into the Schrödinger equation, pre-multiplying by ψ_A, integrating over all space, and then repeating the procedure using ψ_B, we find that the eigenenergies ϵ of the molecular orbitals satisfy

$$\begin{pmatrix} H_{AA}-\epsilon & H_{AB}-\epsilon S_{AB} \\ H_{BA}-\epsilon S_{BA} & H_{BB}-\epsilon \end{pmatrix} \begin{pmatrix} c_A \\ c_B \end{pmatrix} = 0 \tag{2.6}$$

where H_{ij} are the Hamiltonian matrix elements $\langle\psi_i|H|\psi_j\rangle$, and S_{ij} are the *overlap integrals* $\langle\psi_i|\psi_j\rangle$.

Since

$$U_{AB}(\underline{r}) = U(\underline{r}-\underline{R}_A) + U(\underline{r}-\underline{R}_B) \tag{2.7}$$

where $\underline{R}_A$ and $\underline{R}_B$ are the positions of atoms A and B, and U is the potential which enters the Schrödinger equation for the H atom, it follows that

$$H_{AA} = \epsilon_A + \int \rho_A(\underline{r})U(\underline{r}-\underline{R}_B)\,d\underline{r} \tag{2.8}$$

where $\rho_A = |\psi_A|^2$ and ϵ_A is the electron energy for a single H atom. The second term on the right hand side is negative and represents a lowering of the energy of an electron on atom A due to the tail of the potential on the nearby atom B. We will neglect this "crystal-field" shift since it is not essential to understanding the covalent bond. We can also neglect the overlap integrals since these must be small. The key ingredient is H_{AB}, which reduces to

$$H_{AB} = \langle\psi_B|U(\underline{r}-\underline{R}_B)|\psi_A\rangle = H_{BA} \tag{2.9}$$

and is usually called the *bond integral*,[8] and denoted h. It is important to notice that since the hydrogen 1s wavefunction is real and the potential U is negative (i.e. attractive), it follows that $h < 0$.

[7] There is a lot of cheating here. By fixing the positions of the protons we have reduced the three body problem to one involving the movement of a single particle (the electron) in an external field (that produced by the fixed protons).

[8] By inserting Eq. 2.5 into the time dependent Schrödinger equation it can be shown that $|h|$ also gives the transition rate for an electron hopping between the two atoms and hence h is sometimes called the *hopping integral*.

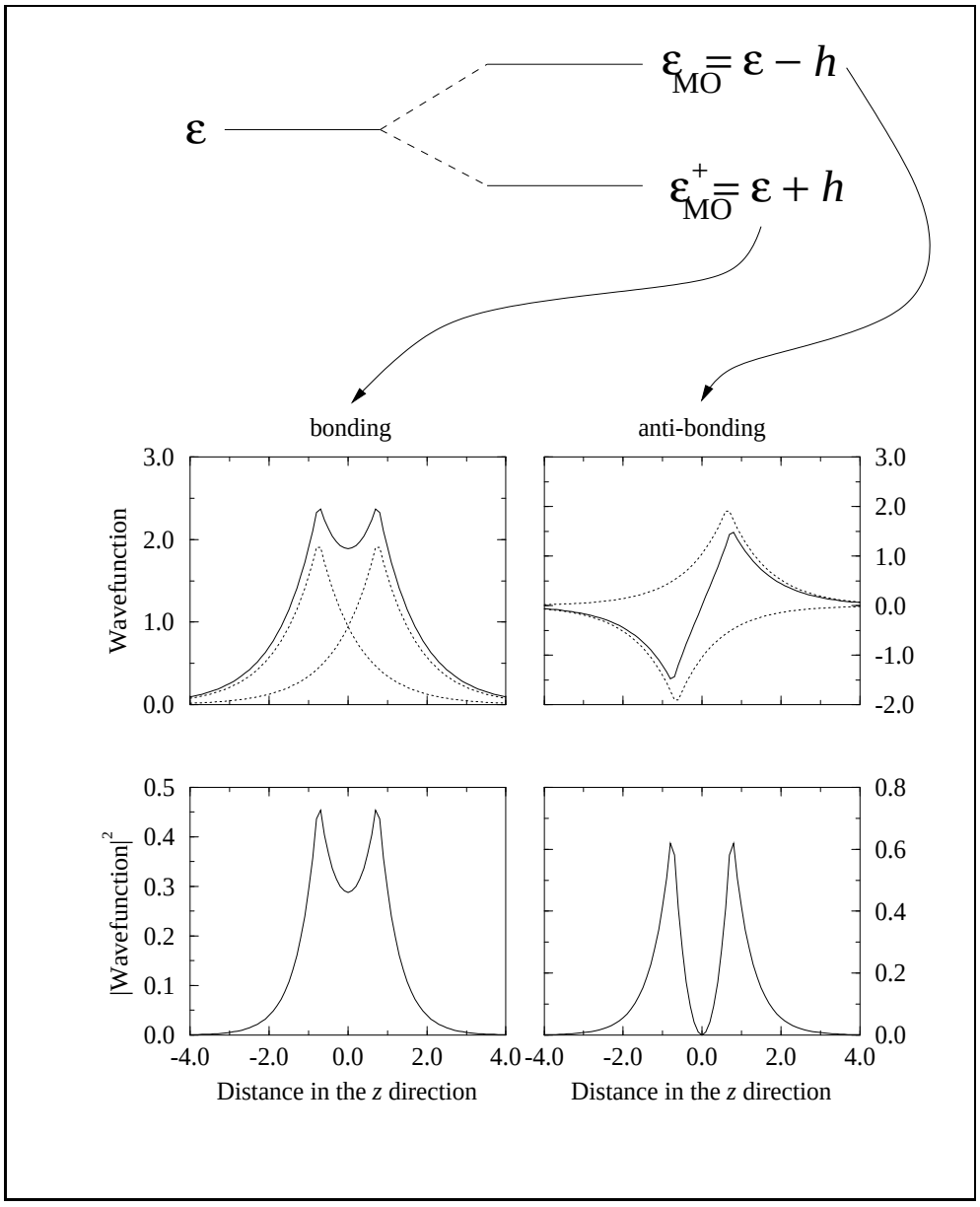

Figure 2.3: Wavefunctions and energies for the H_2^+ molecule.

The MO energies are obtained by setting the determinant of Eq. 2.6 to zero, leading to the two solutions

$$\epsilon_{MO} = \epsilon \pm h. \tag{2.10}$$

Inserting these values back into Eq. 2.6 yields the result $c_A = \pm c_B$ and hence $\psi_{MO} = \psi_A \pm \psi_B$. The H_2^+ wavefunctions and energies are shown schematically in Fig. 2.3. The $c_A = c_B$ solution leads to an increased density in between the protons and a lowering of the energy. We therefore speak of this as the *bonding state* and the internuclear charge as *the bond*. The other state is *antibonding*. If the electron is in this state the molecule will fall apart.

We have seen how combining atomic orbitals can lead to chemical bonding. The simple picture we have presented can be pushed a little further. Imagine now adding a second electron. The Pauli Principle allows us to accommodate this in the bonding orbital. There will be an electron-electron repulsion which would tend to destabilise the molecule a bit but we are entitled to expect the neutral H_2 molecule to hold together. For basically the same reason we would expect Li_2 to be stable. Be_2 is not stable because the Pauli principle would force as many electrons into an antibonding state as there are in the bonding state. There is no net benefit in the molecule forming. Clearly the origin of cohesion in covalent systems is the bonding states getting filled and the antibonding states remaining vacant. $|h|$ is typically a few eV and so covalent bonding is strong.

The covalent Group IV elements are slightly more complicated. As free atoms their valence electron configurations are s^2p^2. Since the s subshell is filled it would appear that only two covalent bonds can be formed. In fact it is highly beneficial to start by promoting the atom to the sp^3 configuration (which costs a small amount of energy) so that four bonds can be formed (which releases a lot of energy). The result when we bring group IV atoms together is tetrahedral coordination. In this case the usual atomic orbitals are not a good starting point for the variational optimization of the molecular orbitals. Instead one starts with "hybridized" orbitals (in this case sp^3 hybrids) which are linear combinations of orbitals on the same site, constructed so as to possess tetrahedral symmetry.

2.1.3 Hydrogen bonding

The inter-molecule bonds in some condensed systems containing hydrogen are stronger than expected. Hydrogen atoms are special because when they participate in a covalent bond with an electronegative atom in a molecule they tend to look like bare protons from the outside. The negatively charged bits of nearby molecules then tend to align themselves so as to maximize the electrostatic attraction with these exposed positive charges. Hydrogen bonding can also occur within big molecules, such as DNA for example where it is responsible for holding the double helix together.

2.1.4 Ionic bonding

To a good approximation, the ionic solid sodium chloride is effectively an array of weakly distorted *ions*, as shown in Fig. 2.1. Although the fluctuating dipole attraction is also present, it is negligible compared to the immense electrostatic

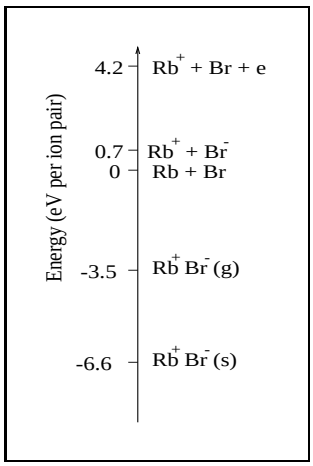

Figure 2.4: Energetics of RbBr.

forces between charged ions.[9] The ions are again prevented from collapsing into each other by Pauli repulsion when the bond length becomes small.

The preference for electron transfer is determined by the energetics of the free atoms. Ionic crystals are formed when strongly electronegative and strongly electropositive ions are combined. This essentially limits the possibility to compounds containing group I, II, VI and VII elements.[10] Let's consider RbBr. It costs 4.2 eV to remove an electron from a Rb atom, i.e. the energy of the Rb atom is 4.2 eV lower than that of the +1 Rb ion. However, the energy of the Br atom is 3.5 eV *higher* than that of the -1 Br ion, and so Br atoms would prefer to have an extra electron. At first sight it would appear that the benefit of transfering an electron to the Br atom would be less than the cost of taking one off the Rb atom. This is only true if the ions are an infinite distance apart. In crystalline RbBr the bond length is 3.4 Å which gives an interaction energy between adjacent ion pairs of $-e^2/4\pi\varepsilon_0 r = -4.2$ eV, which tips the balance strongly in favour of bonding, as shown in Fig. 2.4.

To estimate the cohesive energy we can assume that the dominant contribution is the Coulomb energy of the ion array.[11] This turns out to be a bit more difficult than we might expect, but we are all familiar with the result:

$$E^{coh} = \alpha \frac{q^2}{4\pi\varepsilon_0 r_0}. \tag{2.11}$$

r_0 is the nearest neighbour distance, and $|q|$ is the net charge on each ion. The Madelung constant α is determined by the crystal structure; for the cesium chloride, sodium chloride and zincblende structures α is 1.7627, 1.7476 and 1.6381 respectively. Using experimentally determined lattice parameters and

[9] Typical ionic cohesive energies are of order 8 eV/molecule. The van der Waals interaction is also electrostatic in nature but is only present due to small temporal fluctuations in charge density. The electrostatic forces between charged ions are very strong and time-independent.

[10] Many covalent solids have polar bonds which may be thought of as mixed ionic-covalent bonding.

[11] Since the ions are weakly perturbed we only consider *inter*-ion interactions. To be rigorous we should include the dipolar attraction and the Pauli repulsion but these terms don't change the picture that much.

the appropriate Madelung constant, Eq. 2.11, though a bit crude, predicts the cohesive energies of the alkali halides to within $\sim 10\%$. The CsCl structure has the largest Madelung constant and so has the most favourable Coulomb energy. We are therefore entitled to wonder why all the alkali halides except CsCl, CsBr and CsI adopt the NaCl structure? The answer lies in some simple geometry, as we shall see in SSP.

2.1.5 Metallic bonding

We are familiar with the model of the metallic state in which ion cores are surrounded by a completely delocalized sea of conduction electrons, as shown in Fig. 2.1, but explaining the cohesion of metals is quite involved. We will see later that the free electron model gives a surprisingly good description of the electronic structure of many metals, but it gives rather little insight into cohesion since it accounts only for the electronic kinetic energy. Allowing the electron waves to be delocalized gives a reduction in their average kinetic energy but what is the attractive term which binds the atoms of a metal together? (Experiment tells us that metallic cohesive energies are ~ 1 eV per atom.) As one moves to the left across the periods and down the groups of the periodic table the properties of the elemental solids become more metallic. We would also like to explain this.

One observation is useful however. We know that the maximum of the valence level wavefunctions of the Cu atom, for example, is greater than the interatomic distance in metallic copper. While the valence electrons in metallic copper are not piled up in bonds, it would appear that many neighbouring atoms contribute to the bonding. In fact we might speculate that the metallic state favours maximising the number of nearest neighbours.

2.2 The phases of matter: The simple case of argon

In the previous section we reviewed the basic types of chemical bonding, trying where possible to anticipate the links between bonding and ground state structure. Here we briefly consider what happens for $T > 0$. Again we will focus on the noble gases since for these systems we have an intermolecular potential energy curve (i.e. the Lennard-Jones equation) which is valid for the solid, liquid and gaseous phases.

2.2.1 Classical dynamics?

Just as we used the Lennard-Jones potential to learn about the ground state of argon in §2.1.1, so we should be able to use it to deduce the dynamics of argon atoms (and ultimately the phase diagram), just as we do for the bodies in the solar system using our knowledge of the gravitational force, for example.

• But is it valid to assume that argon atoms can be described by this classical approach?[12]

[12]Note that the ingredients of the Lennard-Jones potential are quantum mechanical in origin. This is obviously true of the Pauli term, but the atomic polarisability and the average

A quick calculation of their thermal wavelength suggests that these atoms will behave as classical particles at any remotely familiar temperature, such as room temperature. But is a classical description of argon atoms valid for $T = 0$? As absolute zero is approached, thermal agitation ceases and one may expect the momenta of the atoms to collapse. In this event the thermal wavelength of the atoms would inevitably exceed the interatomic spacing at low temperture. So do we need a quantum mechanical treatment?[13] No, paradoxically, because of a special quantum mechanical effect stemming from the Heisenberg uncertainty principle ($\Delta x \Delta p \gtrsim \hbar$). This requires that an atom never come to rest since this would imply $\Delta x = \Delta p = 0$. Rather it must retain *zero point motion*. If the noble elements form crystals then Δx must, by definition, be less than the interatomic spacing, and it follows that the de Broglie wavelength is also less than this spacing, and hence a classical description is adequate, provided that we remember that it misses out the zero point energy.

As well as justifying our classical Lennard-Jones approach, we have just stumbled across something useful. The neglected zero point energy is $\sim \frac{\hbar^2}{2M(\Delta x)^2}$ which scales inversely with atomic mass M. Taking Δx to be of the order of Angstroms we find that the zero point energy is just the right order of magnitude to explain the calculated over-binding of the lighter noble solids mentioned above. We should mention here that the zero point energy in He is so great that it prevents a crystal forming even at $T = 0$, unless considerable pressure is applied. At low temperature helium is a very special *quantum fluid*.

2.2.2 The phase diagram

An ensemble of argon atoms may form a solid (high density, incompressible, rigid), a liquid (high density, incompressible, not rigid) or a gas (low density, compressible, not rigid) depending on the temperature T and the pressure P. This behaviour is usually represented using PVT phase diagrams, as in Fig. **??**. You should be familiar with such things from the "Properties of matter" course a couple of years ago, but let's take a quick look.

The boundaries between the solid, liquid and gas phases are shown in the "P-T" diagram on the left. Imagine starting with the gas at temperature T_1 and squeezing it, keeping the temperature constant. We can see what happens by following the T_1 isotherm in the P-V diagram on the right. At first the gas is compressed (i.e. it takes up less volume), then it starts to spontaneously solidify and we enter a region of gas-solid coexistence. In this region the squeezing doesn't increase the pressure but rather it progressivley converts all the matter to the solid phase. When this process is complete we are left with an (almost totally) incompressible solid and so the isotherm turns near vertical. For the T_2 isotherm we see a more elaborate progression: gas to gas/liquid to liquid to

dipole moment are determined by wavefunctions and energies obtained from the hydrogenic Schrödinger equation. We have used quantum mechanical insight to concoct the interatomic potential V, we then imagine the atoms to move like classical particles under the influence of this potential.

[13] Note that this was not the issue that drove us toward a quantum description of the electron gas, since the electrons were assumed to not interact with anything. Here we are assuming the atoms *do* interact, so perhaps we need quantum mechanics to correctly describe the interation between these waves.

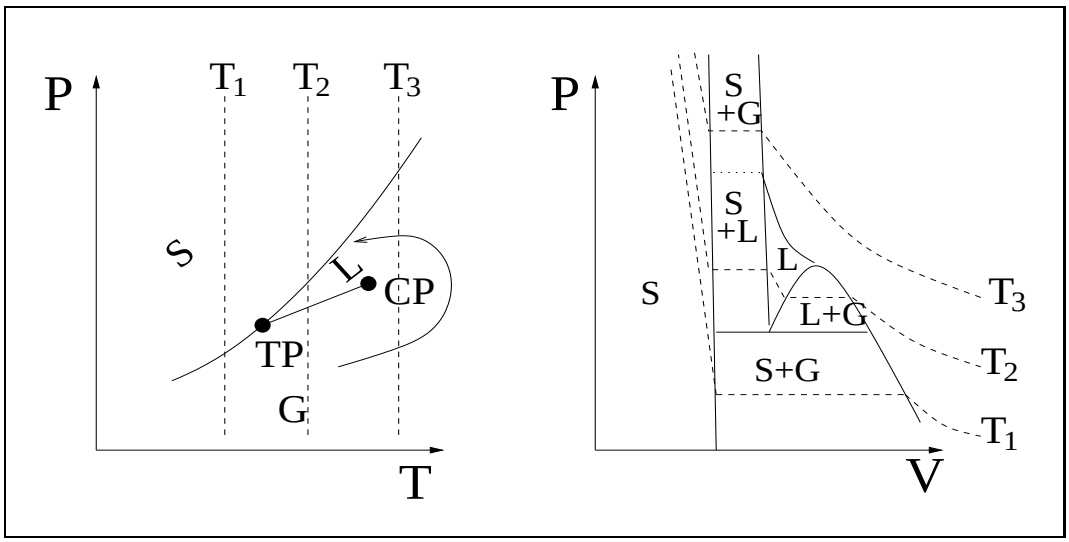

Figure 2.5: Schematic phase diagrams of argon. S means solid, L means liquid and G means gas. TP refers to the "triple point" where all three phases can coexist, while CP is the "critical point" beyong which the liquid-gas phase boundary diappears. The arrow shows that its is possible to go smoothly from the liquid to the gas phase (i.e. without experiencing a phase transtion).

liquid/solid to solid. But then at higher temperature still (i.e. above the *critical temperature* T_c) the liquid phase seems not to be present.[14]

Phase diagrams are determined by making experimental measurements but we also know, in principle at least, how to calculate them: sum up the Boltzmann factors for each configuration of the system to get the partition function, form the Helmholtz free energy, differentiate to obtain the equation of state. This is exceedingly difficult. With some clever approximations, the phase diagram for the Lennard-Jones potential can be computed and such calculations have been able to reproduce experimental results quite well. Here we simply wish to show that a good deal of insight can be gleaned from just the basic facts. We would just like to understand the very basics of condensation (gas to liquid transition) and crystalization (the gas to solid or liquid to solid transition).

At $T = 0$ a system adopts the state for which its energy (in thermodynamics we usually call it the "internal energy" and use the symbol U) is minimized. At finite temperature and with the system in contact with a thermal bath, it is the Helmholtz free energy $H = U - TS$, where S in the entropy, which finds its minimum possible value. If the atoms are far apart then U becomes negligible. It follows that minimization of H is achieved by maximizing the entropy (crudely speaking, the disorder). This is the gaseous state. In forming the liquid and solid states we can see that a trade-off takes place in which a loss of entropy must be outweighed by a lowering of U when the atoms are in close proximity.

The simple notion that *condensation is due to interatomic attraction* appears to be confirmed by the observation that $k_b T_c \sim V_0$, where T_c is the critical

[14]When there is no perceptible gas-liquid phase transition one sometimes describes the system as a *fluid* rather than a gas.

temperature above which the liquid state does not exist. Condensation becomes possible when the average thermal energy becomes less than the interatomic well depth V_0. But why then does the solid form above T_c?

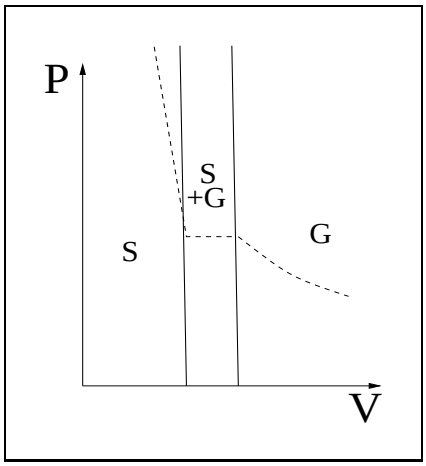

Figure 2.6: Schematic phase diagram of a hard-sphere sysem.

One might wonder what would happen in a system where the constituents don't attract each other at all. The simplest way to explore this is to consider a collection of *hard spheres* for which the interatomic potential consists of just an infinite energy barrier for separations less than an atomic diameter (to prevent overlap) and zero otherwise.[15] One might suppose that T_c would also be negligible for hard spheres and for the phase diagram to simplify somewhat. The phase diagram for such a system is shown in Fig. 2.6. What we see here is pretty much all that we see above T_c for the noble elements. Squeezing a gas of hard spheres leads to steady compression until suddenly crystallization begins when the spheres occupy $\sim 50\%$ of the available volume. There is now insufficient space for individual spheres to migrate around - they are locked into their own place on a crystalline lattice. We can now conclude that *crystallization is due to interatomic repulsion*, which may not be intuitively obvious. Conversion to the solid phase is complete when the occupied volume fraction reaches $\sim 55\%$. This is more or less what happens along a supercritical compression isotherm for real systems like argon, which suggests that we have isolated the crucial physics of crystallization.

There is one more puzzle, however. It is quite hard to explain the freezing of hard spheres from the free energy point of view. Since they don't interact, the system of hard spheres has no potential energy, U is therefore just kinetic energy. At a given temperature this is the same for the solid and the fluid phases. It follows that H can only be lowered by raising the entropy. Intuition would tell us that the solid state would have lower entropy, so how can crystallization occur? The key is to realise that the maximum possible packing of hard spheres would give an occupied volume fraction of 74%, significantly denser than the freezing/melting densities in the hard sphere phase diagram. Thus the balls in the hard sphere solid still have a bit of space in which they can randomly jiggle.

[15] Such systems can be realised in the mesoscopic regime where the constituent is a plastic ball with radius $\sim 10^{-6}$ m.

In fact for density fractions around 50% each ball in the solid is surrounded by more free space than in the hard sphere gas. Jiggling of spheres in their "own space" allows for an increase in entropy upon freezing and hence a lowering in the free energy, as required.[16]

If we squeeze the hard sphere solid we tend to destroy its free local volumes. But we don't produce the zero entropy perfectly close-packed crystal by this method. As we squeeze the jiggling crystal we find that the balls get jammed together in a disordered (finite entropy) fashion at an occupation fraction of 64%. Bernal first demonstrated this *random close packing* phase by pouring small balls into a balloon until it was full, then pouring in glue to fill up the gaps.

2.2.3 Beyond Lennard-Jones?

The simple "6-12" potential is good for argon and the like, but what about other systems? Let's consider just a few.

Liquid crystals

If the molecular subunits are not spherical then the intermolecular potential cannot be "central" (i.e. sperically symmetric). For long, thin and rigid (i.e. rod-like) molecules[17] we get an inetersting complication to the phase diagram. If we cool an isotropic liquid consisting of such molecules a new phase with properties of both a liquid and a solid can be formed. Put crudely, such systems tend to freeze (order) in an anisotropic fashion. The molecular positions remain random but their directions are not. In this *nematic* phase all the long axis of all the molecules tend to point in the same direction, known as the "director". The nematic state is characterised by the ability to flow like a liquid, as the molecules are not locked in position, but also a degree of order, like a crystalline solid, and is therefore referred to as a *liquid crystal* phase.

If a small number of chiral molecules[18] are added to a nematic liquid crystal it is found that the director tilts slightly and also rotates as one moves along its original direction. Depending on the concentration of the chiral component and also the temperature the periodicity of the rotation is about a micron. It follows that these *cholesteric* liquid crystal phases Bragg reflect in the visible region of the electromagnetic spectrum.

Cooling down a nematic liquid crystal may cause the rods to form ordered layers perpendicular to the director. If the molecules are still free to move within these layers, then we have the *smectic* phase.

All this complexity stems from the *shape* of the constituent molecules rather than any elaboration of the underlying physics contained in the Lennard-Jones equation.

[16] We should note that in a real system where interatomic attraction is present the increase in average coordination number (i.e. the average number of nearest neighbours) upon freezing will also tend to lower the Helmholtz free energy of the solid in comparison to the fluid.

[17] Similar effects follow for plate-like molecules.

[18] These are molecules which are structurally inequivalent to their own mirror images. They have right and left handed forms.

Allotropes

Chemical bonding can also greatly complicate phase diagrams. For example, a number of distinct cystalline forms (known as *allotropes*) of the solid state may be observed as the external conditions are varied. The existence of both graphitic and diamond forms of carbon is possibly the most familiar example.

Plastic crystals, quasicrystals ...

While liquid crystals possess orientational order but only limited positional order, the molecules in *plastic crystals* have no orientational order but their centres of mass exhibit translational order.

Condensed matter can exhibit *self-similarity*, which means that it looks the same on all length scales. Such fractal-like structures are ordered since they are determined by a well defined 'rule'. Surprisingly complex (and visually impressive) structures can result from such simple instructions. Examples from everyday life include the fern, snowflakes, the cauliflower and some silica gels. Of course in the real world self-similarity is present over a limited range of length scales.

Some solids, which we call *quasicrystals*, appear to have no periodicity but give rise to diffraction patterns with five-fold symmetry, which was unthinkable until relatively recently.[19] It is has been shown that aperiodic structures in three dimensions can result from taking a 3D "slice" out of a structure which is periodic in a 6D hyperspace. It is not really understood (by me at least) what the physics of this 6D space is.

2.3 Systematic description of atomic structure: Describing disorder

In the previous section we talked qualitatively about the microscopic character of solids, liquids and gases, and about their macroscopic PVT relations. We now start to develop the techniques for systematically describing microscopic structure.

2.3.1 Fluids

If we were able to take a snapshot of an ideal gas and discover the positions of all the atoms we would find them randomly distributed. Knowing that there is an atom at a particular point does not help us predict where any of the other atoms are: we say their positions are *uncorrelated*. Real gasses behave in this way in the high temperature and low density regime. As the temperature is lowered intermolecular attraction starts to become significant as molecules prefer to be near each other. As T is lowered still further these density fluctuations grow in size and live for longer and are more numerous, although the density of the gas remains uniform when the molecular distribution is averaged over a sufficiently

[19] In fact the first experimental observation of five-fold symmetry was met with disbelief and scorn.

large spatial region and time period. Eventually, lowering T will produce a liquid which is distinguished from a gas by its high density.[20]

We tend to think of a liquid as an unstructured ensemble of molecules which are continuously wriggling past each other and there is considerable truth in this picture. However liquids are also distinguished from gasses by virtue of molecular correlation. A schematic representation of a snapshot of liquid argon is shown in Fig. 2.7 and it can be seen that, although long range order is not established, and there also appear to be a few gaps, most argon atoms have an almost close-packed shell of nearest neighbours. Knowing that at a particular instant there is an atom at a particular position allows us to predict with good certainty that there will be approximately 5 or 6 (in 2 dimensions) other atoms about an atomic diameter away. Thus the atomic positions are *strongly correlated* in a liquid.

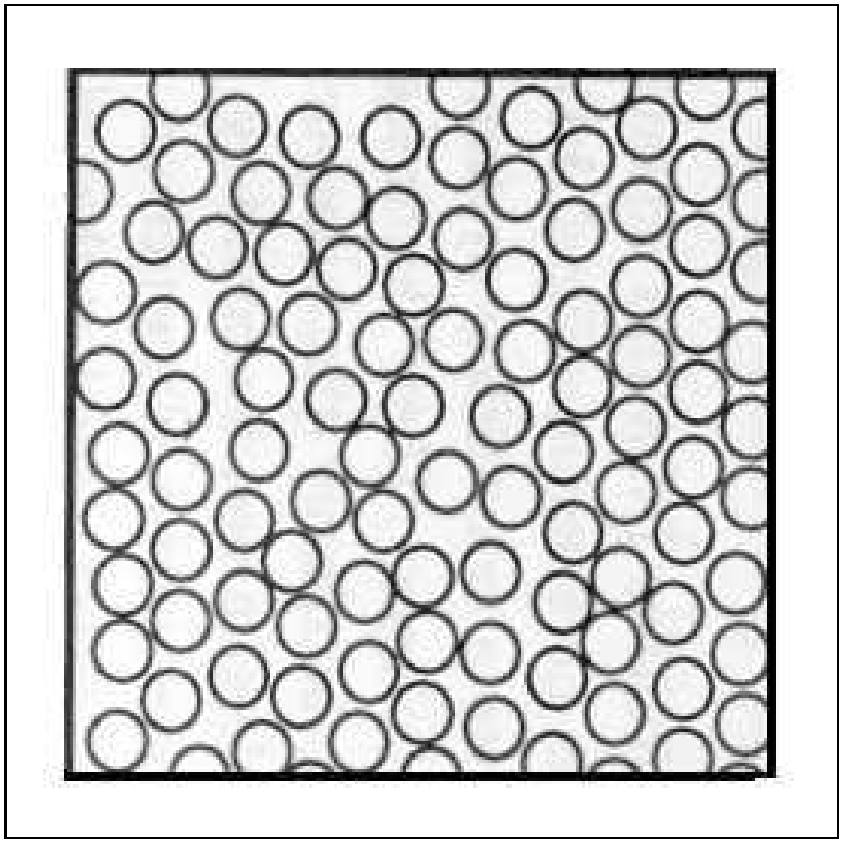

Figure 2.7: Schematic representation of liquid argon.

In order to allow quantification of these fluctuations in density about the global average, we now make a few definitions. The average number density

[20] Strictly speaking we are being a little sloppy here when we make categorical statements about differences between liquids and gases. We already observed in 2.5 that it is possible to convert a liquid to a gas and vice versa without any phase changes taking place. It is only when the two phases *coexist* that we are truly able to distinguish them. In this case there is a visible "meniscus" separating the two phases.

for the whole liquid ρ is just the total number of atoms I divided by the total volume W. If the density varies with position $\underline{r}$, then

$$I = \int \rho(\underline{r})\, d\underline{r} = \int \rho\, g(\underline{r}) d\underline{r}. \tag{2.12}$$

The second equality above defines the *pair correlation function*. This is just the number density in the vicinity of position $\underline{r}$ divided by the average density. To obtain $g(\underline{r})$ we use the expression

$$\rho\, g(\underline{r}) = \langle \sum_i \delta(\underline{r} - \underline{r}_i) \rangle = \frac{1}{I} \sum_{i \neq j} \delta(\underline{r} - \underline{r}_{ij}). \tag{2.13}$$

In essence we choose a particular atom to be at the origin. Then we perform the sum (over all atoms) of delta's in the second term above. It could be that the atom we chose as the origin was not wholly representative of the liquid and so the angle brackets indicate the need to average over all choices of origin. This is done explicitly in the third expression above. Since liquids and gases are isotropic (the same in all directions), $g(\underline{r})$ depends only on r and not on the direction of $\underline{r}$. In this case $g(\underline{r}) = g(r)$ is known as the *radial distribution function* (RDF).

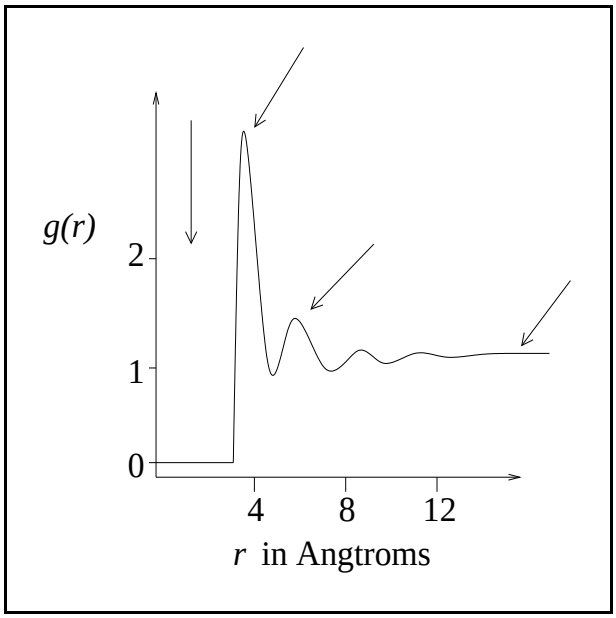

Figure 2.8: The radial distribution function of liquid argon.

To calculate the RDF of argon we can use the Lennard-Jones potential function within a method called "molecular dynamics" as follows. One starts by giving a computer a set of random atomic coordinates and velocities determined by Maxwell-Boltzmann statistics. Given the L-J potential and Newton's second law the computer can then figure out what the atomic positions and velocities will be in a small time later. Recalculating the forces for these new coordinates, the process can be repeated until the average behaviour of the system stabilises. We then have a set of atomic coordinates to insert into Eq. **??**.

Two snapshots of the same place in a liquid taken in rapid succession would be completely different, but the radial distribution function (which is a property of the whole system) would be the same. The simulated RDF for liquid argon near its *triple point* is shown in Fig. 2.8. For small r, $g(r) = 0$ since the atoms cannot overlap. A large peak can be seen near 4 Å, and it is clear that most of the atoms in liquid argon are surrounded by a well defined nearest neighbour "shell". In fact the remnants of several shells can be seen before g converges to a constant value. We say that this liquid, and argon is quite normal in this respect, displays a degree of *short range order*. It is interesting to note that the RDF of the "hard spheres" model (fictitious atoms which have no mutual attraction but infinite repulsion when they overlap), whose phase diagram we saw in Fig. 2.6, is very similar to that shown for argon. This is further confirmation of our deduction that crystallisation (establishment of order) is fundamentally due to interatomic repulsion.

In most liquids the average number of nearest neighbours is slightly less than in the corresponding ordered solid. Thus a decrease in density is observed upon melting. Ice is exceptional in that its volume *decreases* upon melting. This is a consequence of hydrogen bonding. In an ice crystal the water molecules are arranged in a rather open structure. Some of the close-packing favoured by the van-der-Waals interaction is relinquished so that each water molecule can form hydrogen bonds with four nearest neighbours. Upon melting this favoured ordering breaks down, but the coordination number rises to nearer 5 and so the density goes up. As the temperature is increased the normal tendency for nearest neighbour bonds to be lost in the liquid state asserts its authority and above 4 Celsius the density of water diminishes.

Melting is a rather tricky phenomenon to treat in general, but the simple notions of a slightly reduced nearest neighbour co-ordination but a retention of short range order do capture the main essence of the formation of a liquid from a solid.

2.3.2 Amorphous solids

It turns out that the atomic coordinates in some solids is qualitatively similar to what we see for liquids in that they have short range but not long range order. Such solids, illustrated schematically in Fig. 2.9, are known as *amorphous*. The most important distinguishing feature is that two snapshots in rapid succession *would* look the same in amorphous solids, but not liquids. The emphasis here is on the "rapid succession" since the atoms in amorphous solids are mobile if given sufficient time.[21]

Since usually there is a local bonding geometry with a lower energy than competitor geomtries, it follows that *all* atoms will adopt this geometry in the true ground state, giving a perfectly ordered crystal. One might then wonder why amorphous solids exist at all. This is a demonstration of kinetics defeating energetics. The atoms may want to move to a lower energy configuration but they may not have sufficient thermal energy to surmount an activation barrier. Even if there is no barrier the energy gradient driving crystallisation may be rather shallow. It is sometimes useful to think of amorphous solids as extremely viscous liquids.

[21] In fact this is true of the atoms in *any* solid at $T > 0$ as we will see in SSP.

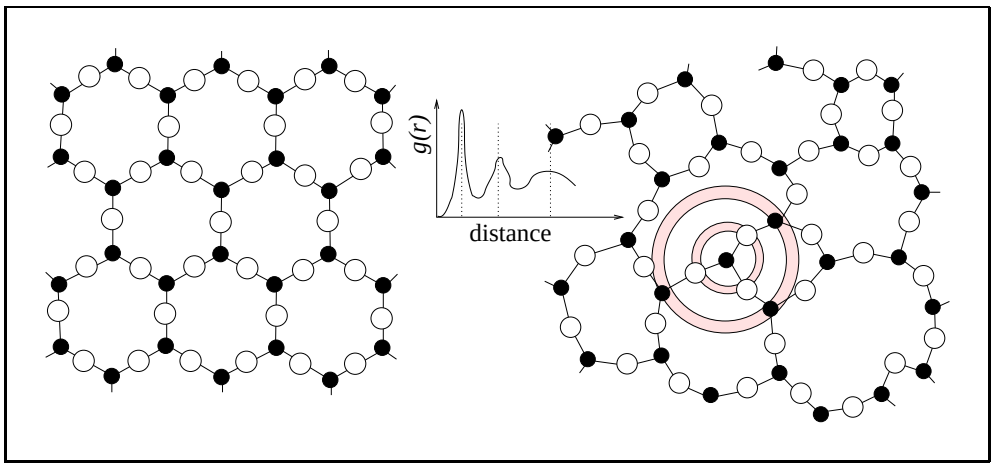

Figure 2.9: Schematic comparison of crystalline and amorphous solids. Also shown is radial distribution function $g(r)$ for the black sites. The well defined coordination shells of the crystal (dotted lines) are smeared out in the amorphous solid.

2.4 Systematic description of atomic structure: Crystallography

In contrast to fluids, crystals are perfectly ordered, with subunits which repeat in all directions. To allow for the most general description of crystals and their properties it is vitally important to make a clear distinction between the subunit and the way that it is repeated:

• A crystal structure is a *Bravais lattice* of points each of which is 'decorated' by a *basis* consisting of one or more atoms.

To make an analogy, we could say that if we were tiling a wall with identical tiles so as to produce a regular pattern, then the basis corresponds to the motif on a single tile, while the Bravais lattice would be the set of positions where a tile is placed.

Before reviewing some of the more common crystal structures we will spend some time considering the Bravais lattices.

2.4.1 General properties of Bravais lattices

A bravais lattice is an infinite array of points which looks the same from whichever of the points the array is viewed.

Alternatively,

A set of points constitutes a Bravais lattice if we can choose any two lattice sites A and B and rigidly translate the entire lattice such that A is now where B used to be, but the lattice looks exactly as it did before.

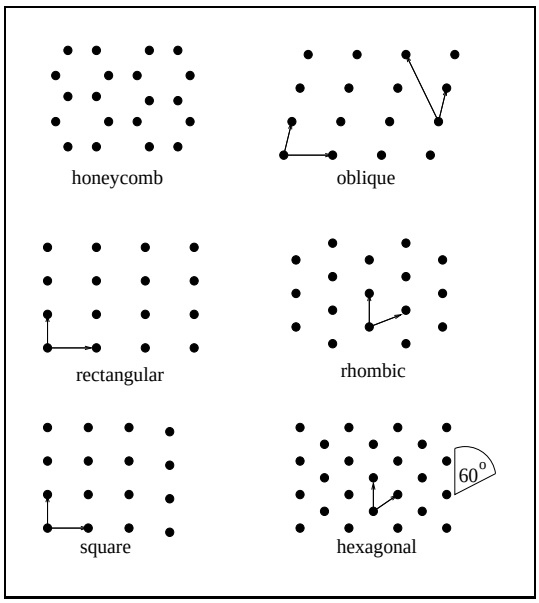

Figure 2.10: A set of six lattices. The dots indicate lattice sites; there are no atoms. All are Bravais lattices except for the honeycomb. Primitive vectors are also shown.

We can cast these rather wordy definitions into mathematical form by stating that the points of a Bravais lattice $\underline{R}$ must satisfy

$$\underline{R} = l\underline{a} + m\underline{b} + n\underline{c} \tag{2.14}$$

where l, m, n take all integer values (positive as well as negative), and $\underline{a}, \underline{b}, \underline{c}$ are three independent vectors which we call *primitive* vectors. We can illustrate the relevant properties of Bravais lattices more easily by recourse to two dimensions.

Two dimensions

Six arrays of points are shown in Fig. 2.10. The honeycomb lattice looks quite regular but is it a Bravais lattice? The answer, which is negative, can be deduced most directly from the second definition above. The others, on the other hand, are all good Bravais lattices. In fact they are the only five Bravais lattices that exist 2D. Oblique is the most general. Square is a special case of either rectangular or rhombic, and hexagonal is a special case of rhombic.

Primitive vectors

For a given Bravais lattice, the primitive vectors are not uniquely defined. Two valid choices are shown for the oblique lattice in Fig. 2.10. Very often the symmetry of a lattice suggests an "obvious" choice however.

Unit cells of a Bravais lattice

A *primitive unit cell* is a volume of space which when translated through all lattice vectors of the form give by Eq. 2.14 exactly fills all space (i.e. there are

no overlaps or gaps). The unit cell corresponds to the *shape of a single tile* in the analogy we introduced above. For a given Bravais lattice there are many (in fact an infinite number of) valid primitive unit cells, each of which contains precisely one lattice site. It is evident that the parallelopiped spanned by three primitive vectors (or the parallelogram spanned by two primitive vectors in 2D) will generate a primitive unit cell.

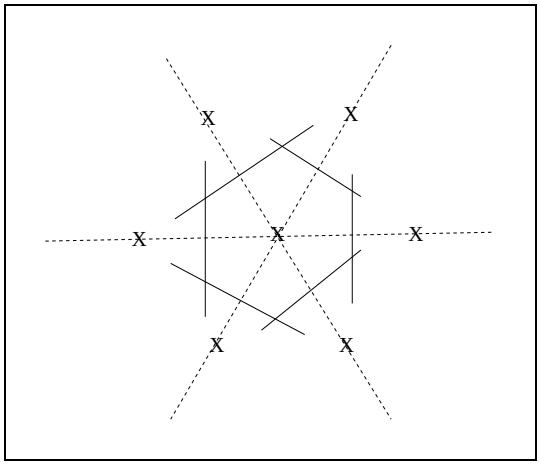

Figure 2.11: Construction of the Wigner-Seitz cell for a 2D Bravais lattice.

For each Bravais lattice a primitive unit cell can be obtained by the *Wigner-Seitz* construction. This is usually defined by (i) drawing lines from one lattice point to all others, (ii) bisecting each of these lines with perpendicular planes, and (iii) taking the smallest polyhedron within these planes. A 2D illustration is shown in Fig. 2.11 All this amounts to is a way of determining the region of space that is closer to the chosen lattice point than to any other lattice point. Since there is nothing in this definition concerning any particular choice of primitive vector, the Wigner-Seitz cell has the full symmetry of the Bravais lattice.

Sometimes it is quite hard to visualise a lattice by looking at a primitive unit cell, and so larger (non-primitive) unit cells are often used. These "conventional unit cells" fill space (when translated through a *subset* of lattice vectors) but contain more than one lattice site. In this way the rhombic lattice shown in Fig. 2.10 is sometimes referred to as the "centred rectangular" lattice and pictured with a rectangular unit cell, as shown in Fig. 2.12. Looking at the solid rectangle, it should be apparent that the rectangular unit cell contains two lattics sites - one in the centre and with the four corner points contributing a quarter each (since they are each shared equally among four idenitcal rectangles). This is perhaps more obviously shown by the dashed rectangle.

Lattice directions

Suppose the components of a Bravais lattice vector are u', v', w' (which are necessarily integers of course) along the $\underline{a}, \underline{b}, \underline{c}$ primitive vectors.[22] If u', v', w'

[22] Note that it is technically necessary to specify which Bravais lattice and which set of primitive vectors are being assumed. For most lattices there is an obvious choice.

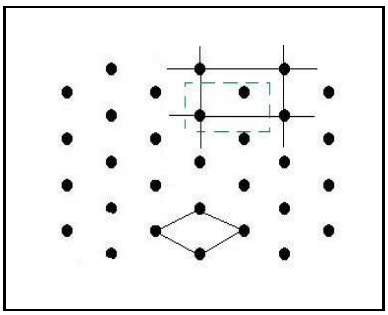

Figure 2.12: Construction of the Wigner-Seitz cell for a 2D Bravais lattice.

are each divided by their highest common denominator to reduce them to the set of smallest integers u, v, w, then the direction of the vector is denoted $[u\,v\,w]$. A complete set of directions equivalent to $[u\,v\,w]$ by symmetry is written $\langle u\,v\,w\rangle$.

Lattice planes

We can decompose any Bravais lattice into a series of parallel planes, each containing the same density and arrangement of lattice points. We can do this in an infinite number of different ways, each producing a distinct set of planes. Our task here is to devise a means of labelling any such set of planes. This can be done by first identifying that plane which passes closest to the origin without actually passing through it. This plane will intersect the axes defined by the primitive primitive cell axes at $u\underline{a}, v\underline{b}$ and $w\underline{c}$, and the reciprocal quantities $1/u, 1/v, 1/w$ will be a set of integers. (Of course u, v and w are not necessarily integers.) We use these integers, which we will call h, k, l, to label the set of planes $(h\,k\,l)$.

We can also decide to find the intercepts u', v', w' of *any* of the planes. In this case we multiply the reciprocal quantities $1/u', 1/v', 1/w'$ by the smallest factor requited to scale them up to integers h, k, l.

The integers h, k, l are called the *Miller indices* of the plane and are dependent on the choice of primitive vectors. A few examples are shown in Fig. 2.13 for the primitive cubic Bravais lattice where we have chosen primitive vectors along the cartesian axes.[23] For this system the direction $[h\,k\,l]$ is then perpendicular to the plane $(h\,k\,l)$, and if the unit cell has length L then the interplanar spacing d_{hkl} is $L/\sqrt{h^2+k^2+l^2}$.

It is possible that certain sets of planes are equivalent due to the symmetry of the Bravais lattice. A complete set of equivalent planes is denoted $\{h\,k\,l\}$.

2.4.2 Bravais lattices three dimensions

There are 14 distinct Bravais lattices in 3D, and crystallographers often group them into 7 *classes* according to their symmetry properties.[24] The Bravais

[23] Note that negative Miller indices are conventionally written with a bar over the number.

[24] A symmetry operation is an operation which when performed on a Bravais lattice leaves the lattice looking exactly the same as it did before. By definition, each Bravais lattice has *translational symmetry*. The set of all symmetry operations is said to constitute the *space group* of a Bravais lattice. Rather than considering all the space group, crystallographers often concern themselves with only those symmetry operations which leave one point of the

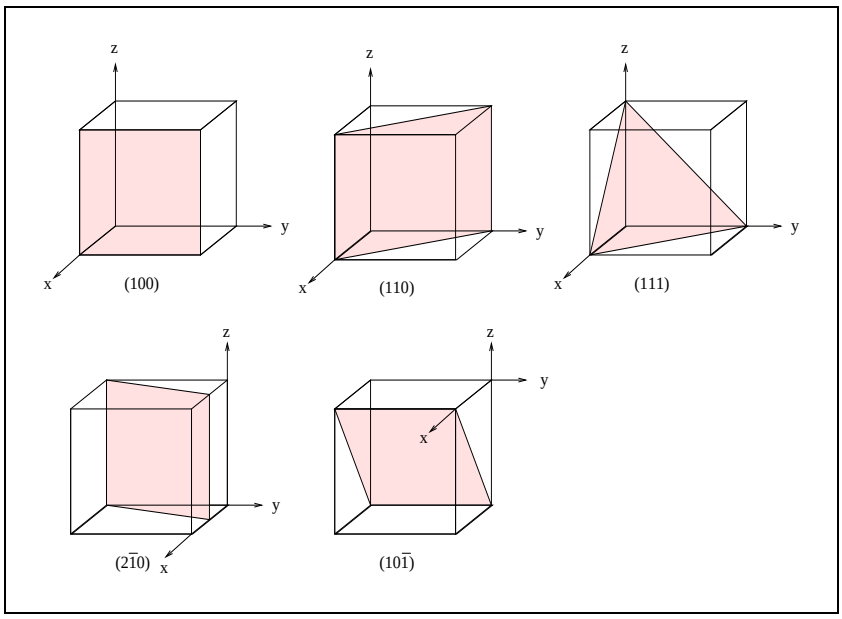

Figure 2.13: Lattice planes for the "primitive cubic" Bravais lattice.

lattices are enumerated in most reasonably advanced texts on the physics of solids, but we will only need to be familiar with a few of them.

The primitive cubic (or simple cubic) Bravais lattice

This can be generated[25] by the primitive vectors $L\underline{\hat{x}}, L\underline{\hat{y}}, L\underline{\hat{z}}$. The conventional unit cell is the cube of side L, as is the Wigner-Seitz cell in this case.

The body centred cubic Bravais lattice

The conventional unit cell for this Bravais lattice is shown in Fig. 2.14. It is a cube of side L with lattice points at the eight corners and one at the centre of the cube. Clearly this is a non-primitive unit cell since it contains 2 lattice sites (one in the middle and eight corners sites each with weight 1/8), but drawing it this way helps us understand the relationship between the lattice sites. The Wigner-Seitz cell for this lattice, a truncated octahedron, is also shown. The body centred cubic Bravais lattice can be generated by the primitive vectors $(\underline{\hat{x}} + \underline{\hat{y}} - \underline{\hat{z}})L/2, (\underline{\hat{x}} - \underline{\hat{y}} + \underline{\hat{z}})L/2, (-\underline{\hat{x}} + \underline{\hat{y}} + \underline{\hat{z}})L/2$.

The face centred cubic Bravais lattice

The face centred cubic Bravais lattice is like the primitive cubic but with an extra lattice site at the centre of every face of the cubes, and it can be generated

lattice fixed. There are only 7 distinct *point groups* that a Bravais lattice can have, and so it is conventional to divide them into 7 *crystal systems*. In this course we will only exploit translational symmetry, and so we don't need to get into group theory.

[25] There is no unique choice of primitive vectors for a given Bravais lattice. But a particular set of primitive vectors generates a unique Bravais lattice.

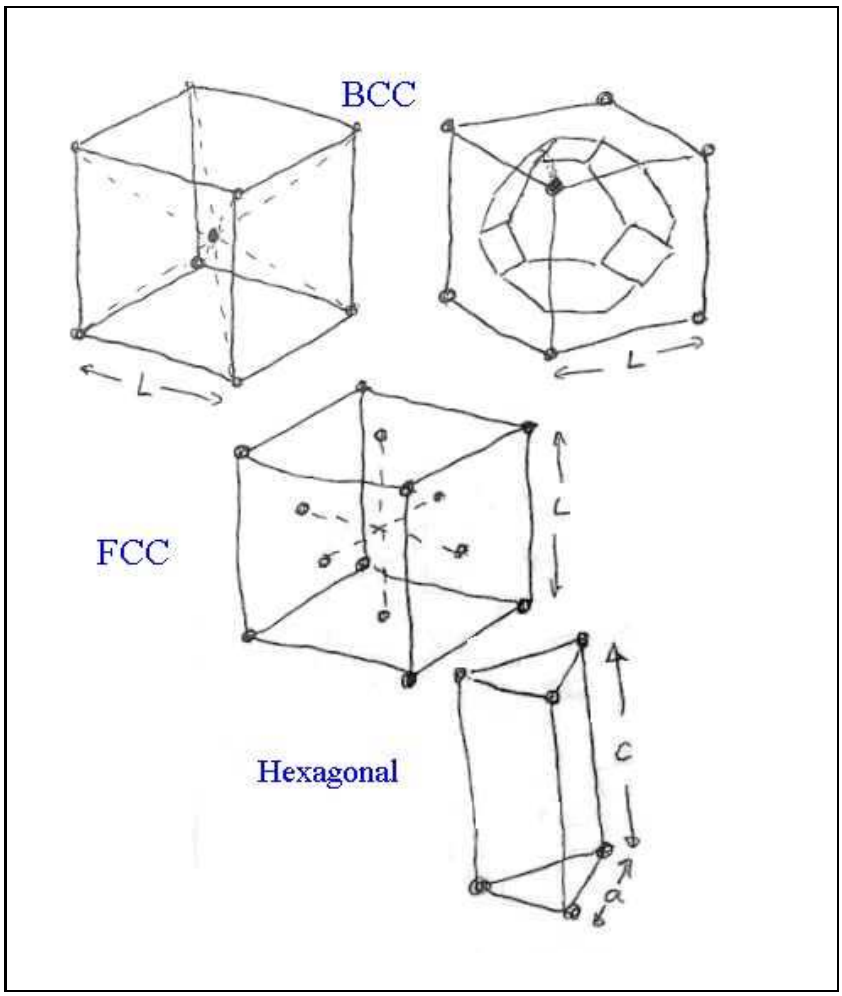

Figure 2.14: Unit cells for the body centred cubic (top), face centred cubic (middle), and simple hexagonal (bottom) Bravais lattices. The Wigner-Seitz unit cell for the BCC Bravais lattice is also shown.

by the primitive vectors $(\underline{\hat{x}} + \underline{\hat{y}})L/2, (\underline{\hat{y}} + \underline{\hat{z}})L/2, (\underline{\hat{z}} + \underline{\hat{x}})L/2$. The conventional unit cell for this lattice contains four sites (8 corner sites, each worth 1/8, plus 6 face sites, each worth 1/2) and is shown in Fig. 2.14. The Wigner-Seitz cell is a rhombic dodecahedron and is a bit tricky to draw.

The simple hexagonal Bravais lattice

This lattice is generated by the primitive vectors $\underline{\hat{x}}a, \underline{\hat{x}}a/2 + \underline{\hat{y}}\sqrt{3}a/2, \underline{\hat{z}}c$. The primitive cell is a "toblerone" with length c and triangle of side a, again shown in Fig. 2.14.

2.4.3 Some common crystal structures

Before some real crystal structures we will again drop down to two dimensions just to get the hang of things. It's quite simple really: "crystal = Bravais lattice with a basis".

The basis

Up to this point we have only been speaking of lattice sites. A crystal was defined above as a Bravais lattice with each lattice site given an identical *basis* of atoms. The basis is simply a group of atoms (possibly only one atom) with a definite spatial relationship. When there is more than a single atom in the basis we must clearly specify *what and where*: the atomic number or element of each of the atoms, together with their coordinates. It is often convenient to give atomic coordinates in terms of fractions of the conventional unit cell, known as *fractional coordinates*.

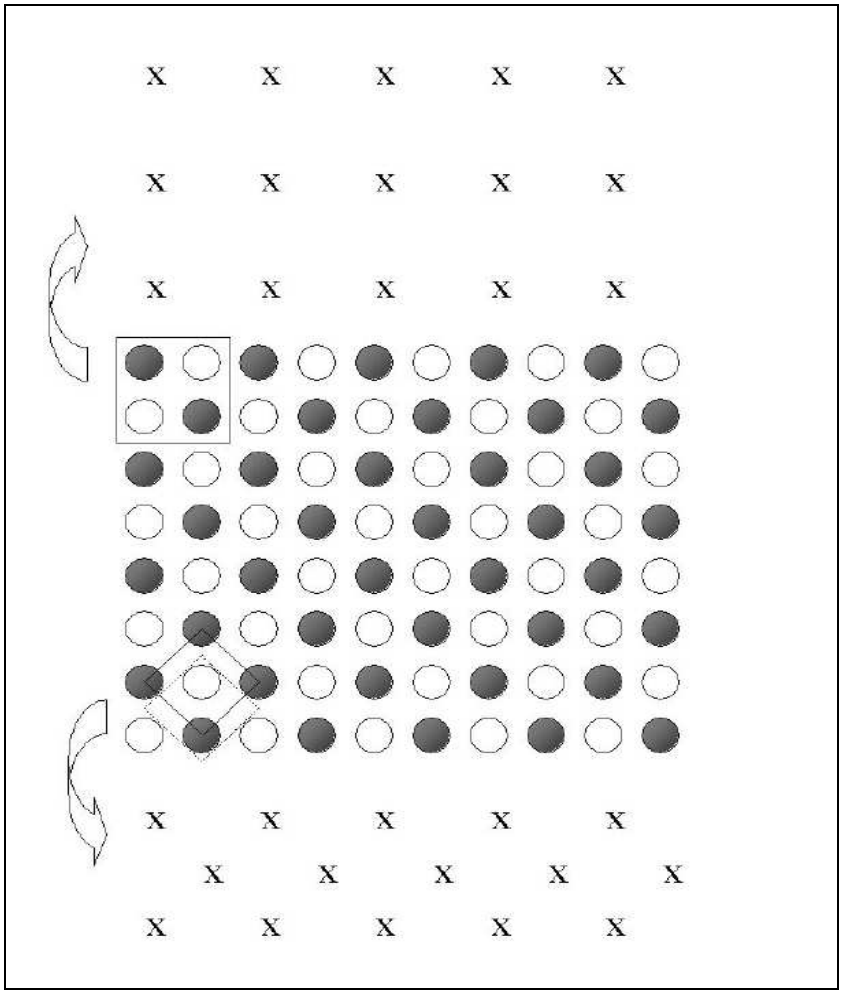

Figure 2.15: A 2D crystal comprising black and white atoms.

Fig. 2.15 shows a 2D crystal comprising black and white atoms. Quite obviously it is not possible to describe such a crystal using a single atom basis; we need a basis with at least one black and one white atom. Two possibilities are shown in the figure, the upper one comprising a four atom basis and the lower one a two atom basis. In each case the corresponding Bravais lattice is also shown. While we note that the expression of a crystal structure as a particular Bravais lattice with a particular basis is not unique, we will see that the most compact and simplest description is the one with the smallest possible basis. An apparent shortcoming of the small unit cell in the lower left hand corner is that it contains fractions of atoms. Rather than try to keep track of all these bits it is easier to imagine tugging the unit cell down a little, as shown by the dotted square, so that it contains two complete atoms. We can then specify the basis as: a black atom at (0,0) and a white atom at (0.5,0.5).

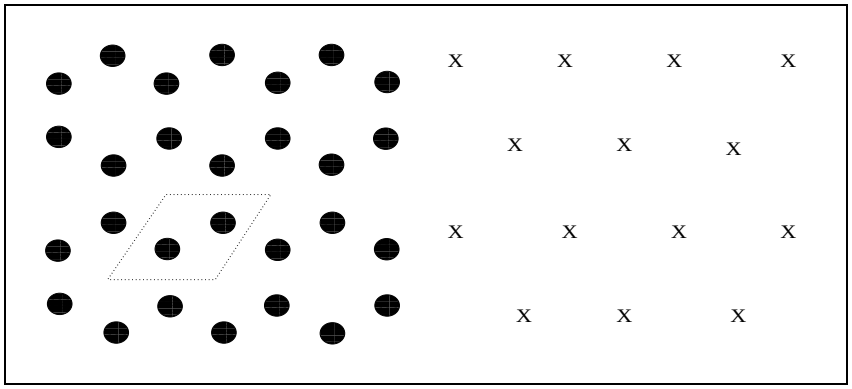

Figure 2.16: The 2D honeycomb structure is shown on the left. This can be viewed as a hexagonal Bravais lattice, shown on the right, together with the two atom unit cell indicated by the dotted lines.

It is less obvious that a multiatom basis is required when all the atoms are of the same type, such as the example in Fig. 2.16. Although each atom (shown as a black circle) has an identical bonding environment, the atomic sites do not constitute a Bravais lattice since they do not satisfy Eq. 2.14, as we observed when considering 2.10 on page 35. We should view the honeycomb crystal as a hexagonal Bravais lattice, shown on the right of Fig. 2.16, with a two atom basis. We will see that a number of the common 3D crystal structures must similarly be described with a multi-atom basis despite their apparent simplicity.

The simple cubic crystal structure

Contrary to the billing of this section, this is an exceedingly rare crystal structure in nature with only one element (polonium) adopting it. The simple cubic structure shown in Fig. 2.17. It should be obvious that this structure can simply be described as a primitive cubic Bravais lattice with spacing L and a single atom at each lattice site. There is no point in doing so, but we could also regard this structure as a primitive cubic Bravais lattice with spacing $2L$ together with an eight atom basis with atoms at fractional coordinates $(0,0,0)$, $(0,1/2,0)$, $(0,0,1/2)$, $(1/2,0,0)$, $(1/2,1/2,0)$, $(1/2,0,1/2)$, $(0,1/2,1/2)$ and $(1/2,1/2,1/2)$.

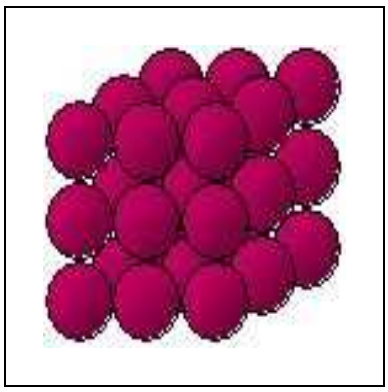

Figure 2.17: The simple cubic crystal structure.

Alternatively, we could describe it as a face centred cubic Bravais lattice with cube dimension $2L$, together with a basis with atoms at $(0,0,0)$ and $(1/2,0,0)$.

The body centred cubic crystal structure

The simplest description of this structure is as a body centred cubic Bravais lattice with a one atom basis.

The face centred cubic crystal structure

The simplest description of this structure is as a face centred cubic Bravais lattice with a one atom basis.

The diamond crystal structure

This is the structure adopted by diamond, silicon, germanium and one form of tin, and shown in Fig. 2.18. We can see that each atom is bonded to four others and they occupy chemically equivalent environments. But do the atoms fall on the sites of a Bravais lattice? No, this is the 3D equivalent of the honeycomb structure we saw earlier. This time we need to specify a multi-atom basis *by necessity*.

It's rather hard to really understand the relative positions of the atoms by just looking at the conventional unit cell (the cube). To make it a little easier the conventional unit cell is also shown in the figure as a series of slices. You may now be able to appreciate that the diamond crystal structure can be expressed as an FCC Bravais lattice and a basis consisting of an atom at $(0,0,0)$ and another at $(1/4,1/4,1/4)$, where the fractional coordinates are in terms of the conventional unit cell.

Packing fraction and coordination number

Earlier in this chapter we reflected on the fact that the noble elements should probably form crystal structures in which the atoms as densely packed as possible. If we were to imagine the atoms in the four crystal structures considered so far to be spheres packed as closely as the structure will allow, as shown in Fig. 2.19, then a little geometry would reveal that the fraction of space occupied by

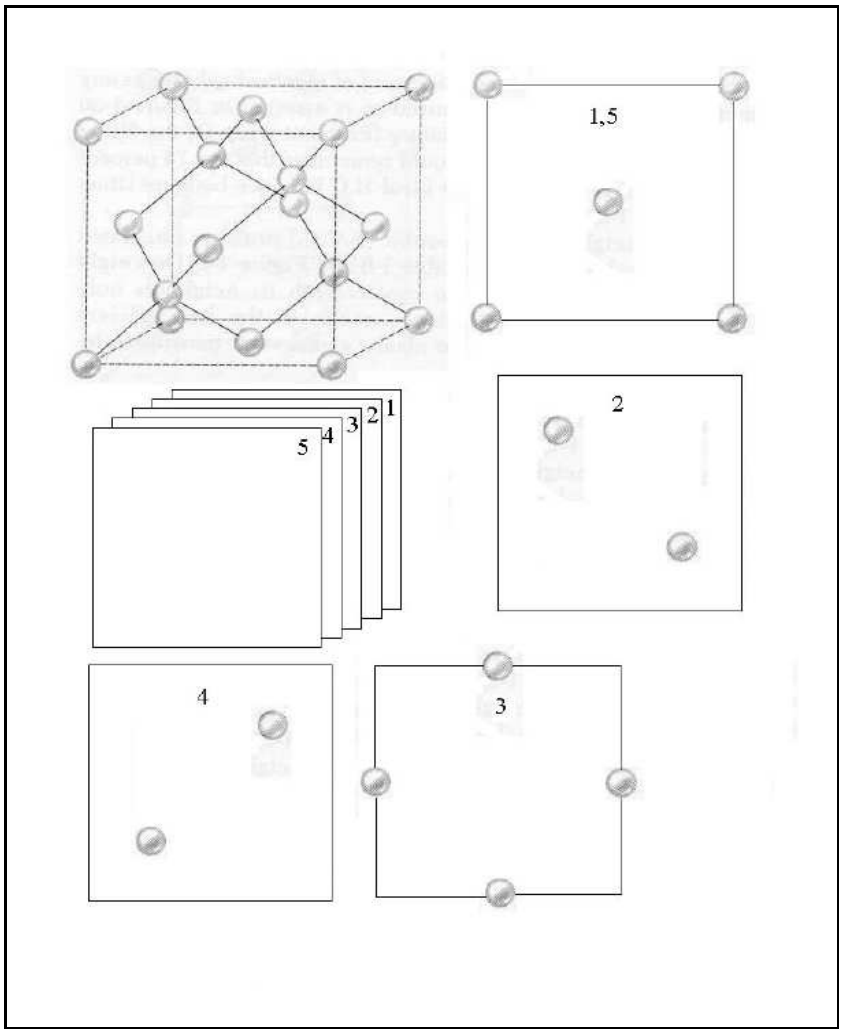

Figure 2.18: The tetrahedral bonding arrangement in the diamond structure is shown on the left. On the right, the conventional unit cell is shown.

the atoms would be 0.52, 0.68, 0.74 and 0.34 for SC, BCC, FCC and diamond crystal structures respectively. Thus the noble elements, which are observed to crystallise with the FCC structure do indeed appear to have maximized their packing, each atom surrounded by 12 nearest neighbours.

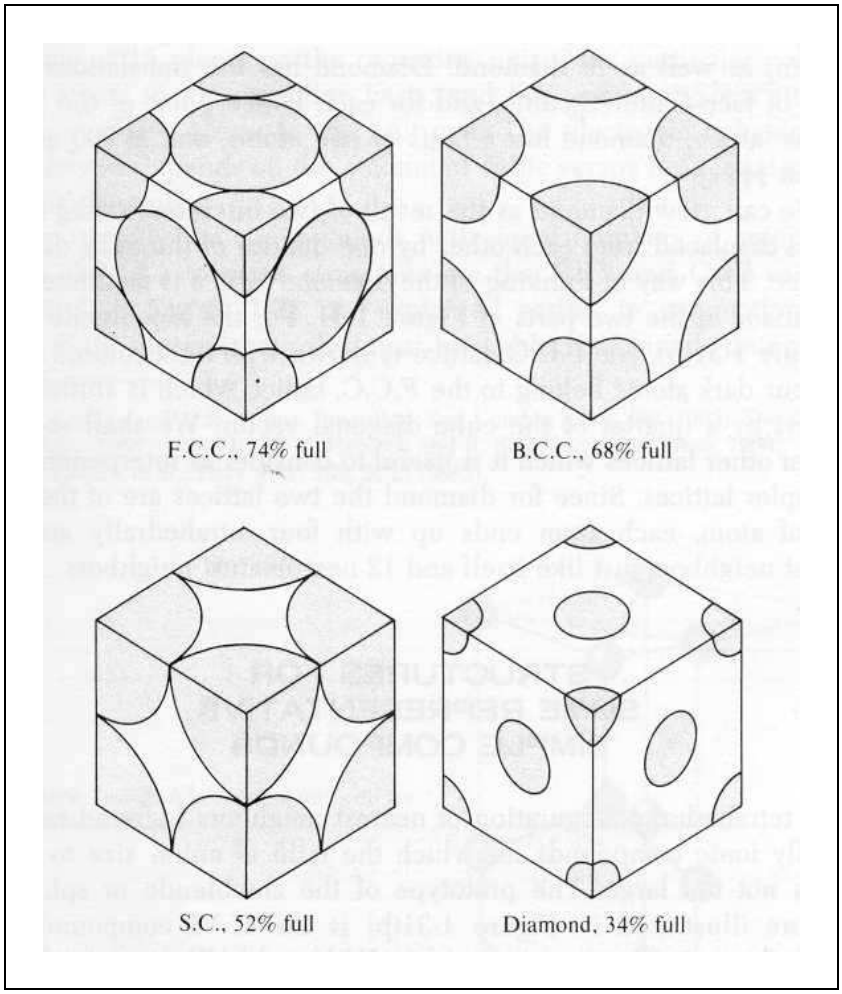

Figure 2.19: The conventional unit cells of the simple cubic, face centred cubic, body centred cubic and tetrahedral structures drawn to illustrate packing density.

The FCC structure is said to be *close-packed*. Although it may not be immediately obvious, the atoms are as close together as geometry allows, as Fig. 2.20 attempts to illustrate. Here a 2D plane of densely packed atoms are revealed. The FCC structure can be viewed as a series of such planes of atoms stacked on top of each other with the atoms of one plane nestling in the hollows of the adjacent planes. A little thought reveals that any two planes can be fitted together in two ways - if the lower plane is "A", then the upper plane can be "B" or "C" type, as shown in Fig. 2.21. With reference to Fig. 2.20 we see that the FCC structure is "ABCABC...".

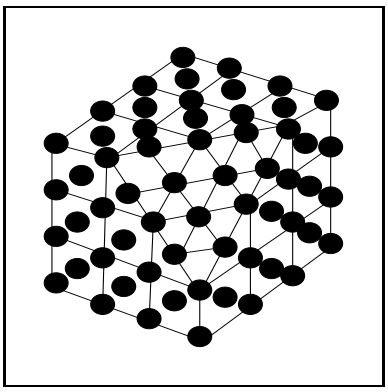

Figure 2.20: Close-packing in the FCC structure.

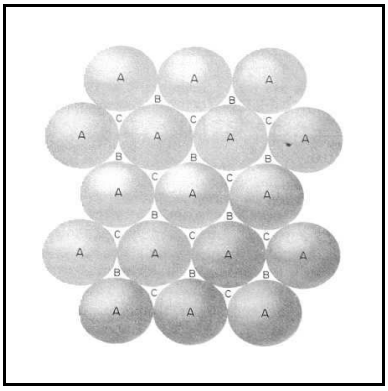

Figure 2.21: Different ways of stacking of close-packed planes.

What about metals? Many (such as Pd, Cu, Al, Ni, Ag, Au, Rh, Pb) are FCC and many are BCC (such as Li, Na, K, Rb, Cs, Ba, Fe, V, W, Ta) at room temperature, and quite a few switch as T is varied. It seems that close-packing is favoured but a delicate balance exists between 12-fold coordination (FCC) and 8-fold (BCC). The second nearest neighbours in the BCC structure are not that further distant than the nearest neighbours so this makes sense.

The hexagonal close-packed crystal structure

Returning to the stacking of close-packed layers, it is clear that any sequence of A, B and C will lead to close-packing. ABABAB... is what we call *hexagonal close-packing*. It is easily shown that the atoms in this structure do not lie on a Bravais lattice - again it is like the honeycomb situation. The simplest description is as a hexagonal Bravais lattice with a basis of atoms at $(0, 0, 0)$ and $(1/3, 1/3, 1/2)$. For true close-packing the c/a ratio (see Fig. 2.14) equals 1.63 but this can vary without changing the Bravais lattice. About thirty of the elements adopt this structure under normal conditions.

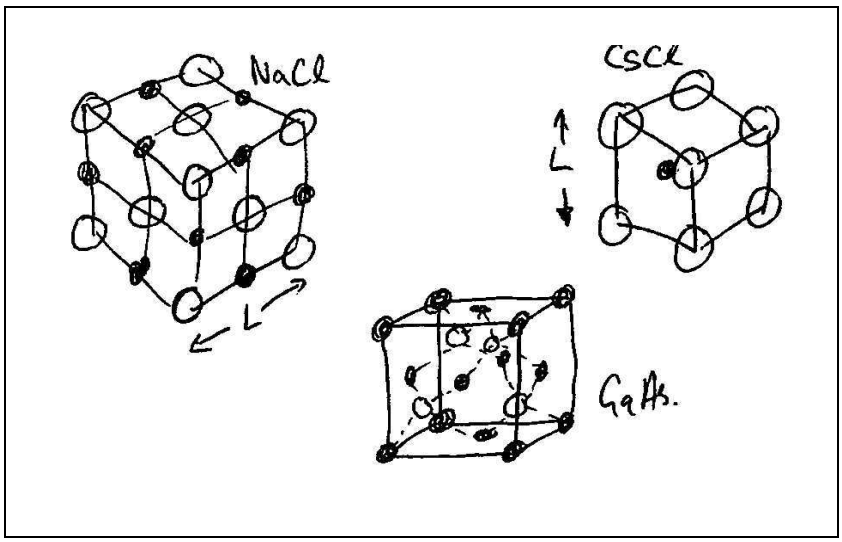

Figure 2.22: The GaAs, NaCl and CsCl structures.

The gallium arsenide crystal structure

This structure shown in Fig. 2.22 is like diamond but with half the atoms Ga and half As. To describe this structure we simply need to repeat the description of diamond but state that the atom at $(0,0,0)$ is Ga, and that at $(1/4,1/4,1/4)$ is As.

The sodium chloride crystal structure

Here it is easy to mess up because this looks so much like the simple cubic structure. But think: what is the Bravais lattice and what is the basis? Clearly we need at least a two atom basis. But notice that the Na ions are arranged in an FCC structure, and so are the Cl ions. So we can describe this structure as an FCC Bravais lattice with conventional cube dimension L, together with a basis with an Na ion at $(0,0,0)$ and a Cl ion at $(1/2,0,0)$.

The Cesium chloride crystal structure

Again it is easy to be seduced here, this time by the similarity with the BCC crystal structure. In fact we have a simple cubic Bravais lattice with a basis comprising a Cs ion and a Cl ion.

2.5 The reciprocal lattice

In the previous section we introduced the notion that a crystal structure is defined by a Bravais lattice convoluted with a basis. This all seems a bit heavy handed at first, but it is worthwhile in the long run since this scheme offers the most compact description of crystals that can be devised. We need only specify the contents of the basis and instructions for how to duplicate this throughout all space.

In this section we will add to the formal machinery by introducing the *reciprocal lattice*. We will see that this concept provides an extremely compact description of diffraction, but it gives us the appropriate framework to handle waves of any kind within a crystal. In passing, we mentioned in Chapter 1 that for every symmetry there is a corresponding conservation law. In a crystal we do not have translational invariance, but translational periodicity. We will see (mainly in SSP next term) that in such circumstances the conservation of electron momentum, witnessed in the free electron gas, is replaced by a more subtle conservation law in which the reciprocal lattice plays a role. Given your expertise with Fourier's theorem, it will come as no surprise that reciprocal lattice vectors enter the equations when we consider basically any property of crystals since they are, by definition, periodic and hence so are their properties. Let's start with a quick recap and a definition.

2.5.1 Periodicity and the Fourier expansion

Any function which is periodic:

$$f(x) = f(x + L) \tag{2.15}$$

can be expressed as a Fourier series:

$$f(x) = \sum_{n=-\infty}^{+\infty} f_n e^{in2\pi x/L} \tag{2.16}$$

where n can be any integer, and the Fourier coefficients f_n are given by

$$f_n = \frac{1}{L}\int_0^L f(x)e^{-in2\pi x/L}\,dx. \tag{2.17}$$

Alternatively, we could write Eq. 2.16 as

$$f(x) = \sum_K f_K e^{iKx}, \qquad K = \frac{2\pi}{L} \times n. \tag{2.18}$$

The Fourier expansion works because the complex exponentials are also periodic with period L, and they are a complete orthonormal set of functions.

Moving to 3D things are very similar. Any periodic function

$$f(\underline{r}) = f(\underline{r} + \underline{R}) \tag{2.19}$$

where $\underline{R} = u\underline{a} + v\underline{b} + w\underline{c}$ with u, v and w integers (i.e. $\underline{R}$ is a set of Bravais lattice vectors) can be expressed as a Fourier series:

$$f(\underline{r}) = \sum_{\underline{K}} f_{\underline{K}} e^{i\underline{K}\cdot\underline{r}}. \tag{2.20}$$

But what are the vectors $\underline{K}$ in this case?

As in the 1D case, these must have the periodicity of the $\underline{R}$ lattice.[26] The set of all vectors $\underline{K}$ which correspond to plane waves with wavevectors $\underline{K}$ having the

[26] Note that in Chapter 1 we introduced periodic boundary conditions and we saw that this meant that the allowed vectors in the free electron gas are quantised. This was just an artificial device to allow us to count the states. The period in this case is extremely large, macroscopic in fact. Here we are talking about the periodicity from unit cell to unit cell in a crystal. The period is of the order of atomic dimensions.

periodicity of a given $\underline{R}$ lattice is known as its *reciprocal lattice*. Mathematically this means

$$e^{i\underline{K}.(\underline{r}+\underline{R})} = e^{i\underline{K}.\underline{r}} \tag{2.21}$$

or simply

$$e^{i\underline{K}.\underline{R}} = 1 \qquad \Rightarrow \qquad \underline{K}.\underline{R} = 2\pi \times (\text{integer}) \tag{2.22}$$

for any $\underline{R}$ in the Bravais lattice. In other words, a reciprocal lattice vector is the wavevector of a plane wave that has the same value at all points of the direct Bravais lattice. In fact, a plane wave takes the same value on a plane perpendicular to its wavevector (and at all such planes separated by an integral number of wavelengths), alerting us to the intimate relationship between reciprocal lattice vectors and planes of lattice sites in the direct lattice.

Having defined a "reciprocal" lattice, one sometimes refers to the original as the "direct" lattice.

2.5.2 Primitive vectors

If $\underline{a}, \underline{b}, \underline{c}$ are primitive vectors of a lattice in real space (the *direct* lattice), the vectors defined by

$$\begin{aligned}
\underline{a}^* &= 2\pi \frac{\underline{b} \times \underline{c}}{\underline{a}.(\underline{b} \times \underline{c})} = \frac{2\pi}{V} \underline{b} \times \underline{c} \\
\underline{b}^* &= 2\pi \frac{\underline{c} \times \underline{a}}{\underline{b}.(\underline{c} \times \underline{a})} = \frac{2\pi}{V} \underline{c} \times \underline{a} \\
\underline{c}^* &= 2\pi \frac{\underline{a} \times \underline{b}}{\underline{c}.(\underline{a} \times \underline{b})} = \frac{2\pi}{V} \underline{a} \times \underline{b}
\end{aligned} \tag{2.23}$$

where V is the volume of the unit cell of the direct lattice, are primitive vectors of the reciprocal lattice. The reciprocal lattice vectors have the form

$$\underline{K} = h\underline{a}^* + k\underline{b}^* + l\underline{c}^* \tag{2.24}$$

where h, k and l are integers.

Eq. 2.23 is very useful in 3D. An alternative definition which also works in different dimensional spaces is

$$\underline{\alpha}.\underline{\beta}^* = 2\pi\delta_{\alpha\beta} \tag{2.25}$$

where α and β can each be $\underline{a}, \underline{b}$ or $\underline{c}$.

It is important to appreciate that although the primitive reciprocal lattice vectors one obtains depend on the particular choice of primitive vectors $\underline{a}$, $\underline{b}$, $\underline{c}$, the reciprocal lattice they generate (Eq. 2.24) does not. This must be so since no such choice enters the definition of the reciprocal lattice given by Eq. 2.22.

2.5.3 Properties

1. The reciprocal lattice is itself a Bravais lattice (and so one can construct *its* reciprocal).

2. The reciprocal of the reciprocal lattice is just the original direct lattice.

3. If V is the volume of the unit cell of the direct lattice, the volume of the unit cell of the reciprocal lattice is $V^* = (2\pi)^3/V$.

4. Reciprocal lattice vectors have dimensions of "inverse length", or $(\text{length})^{-1}$ (hence the name).

2.5.4 Examples

1D

In one dimension vector notation becomes redundant. The direct Bravais lattice "vectors" are uL, where u is the set of all integers and L defines the length scale. The corresponding reciprocal lattice "vectors" are just $h\frac{2\pi}{L}$, where h is the set of integers. Notice that wide spacing in real space (i.e. large L) leads to narrow spacing in reciprocal space.

2D

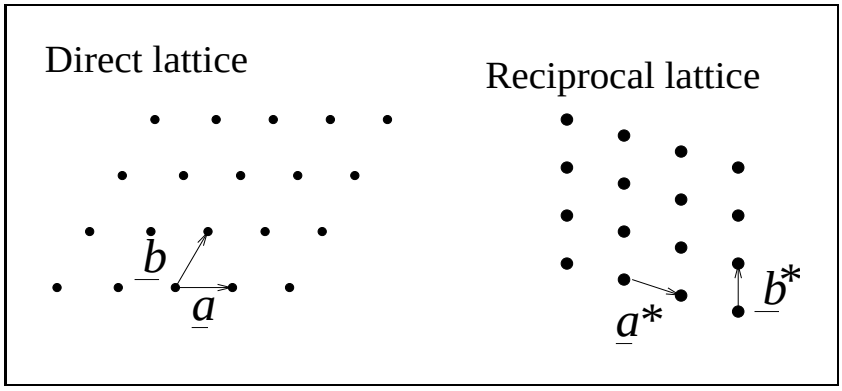

Figure 2.23: A 2D Bravais lattice and its reciprocal lattice. Primitive vectors are indicated in each case.

A 2D example is shown in Fig. 2.23. Notice that $\underline{a}^*$ is perpendicular to $\underline{b}$, and $\underline{b}^*$ is perpendicular to $\underline{a}$. This means that $\underline{a}^*$ is in general not perpendicular to $\underline{a}$.

Simple cubic direct Bravais lattice

The obvious choice of primitive vectors for the simple cubic crystal structure are $\underline{a} = L\,\underline{\hat{x}}, \underline{b} = L\,\underline{\hat{y}}$ and $\underline{z} = L\,\underline{\hat{z}}$, where L is the lattice constant (the length of the cube sides), and $\underline{\hat{x}}, \underline{\hat{y}}, \underline{\hat{z}}$ are unit cartesian axis vectors. Eq. 2.23 yields $\underline{a}^* = 2\pi/L\,\underline{\hat{x}}, \underline{b}^* = 2\pi/L\,\underline{\hat{y}}$ and $\underline{c}^* = 2\pi/L\,\underline{\hat{z}}$, and so the reciprocal lattice of the

simple cubic direct lattice of side L is just another simple cubic lattice but with side $2\pi/L$.

Face-centred cubic and body-centred cubic direct Bravais lattices

Primitive vectors for the FCC lattice are provided by $\underline{a} = L/2\,(\hat{\underline{y}} + \hat{\underline{z}})$, $\underline{b} = L/2\,(\hat{\underline{z}} + \hat{\underline{x}})$, and $\underline{c} = L/2\,(\hat{\underline{x}} + \hat{\underline{y}})$, where L is the lattice constant (defined to be the length of the sides of the conventional FCC unit cell). Applying Eq. 2.23 we find $\underline{a}^* = 2\pi/L\,(-\hat{\underline{x}} + \hat{\underline{y}} + \hat{\underline{z}})$, $\underline{b}^* = 2\pi/L\,(\hat{\underline{x}} - \hat{\underline{y}} + \hat{\underline{z}})$, and $\underline{c}^* = 2\pi/L\,(\hat{\underline{x}} + \hat{\underline{y}} - \hat{\underline{z}})$. These are primitive vectors for a BCC lattice with lattice constant (defined as the length of the sides of the conventional BCC unit cell) $4\pi/L$. Thus the reciprocal of the FCC lattice is a BCC lattice, and vice versa.

2.5.5 Brillouin zones

The Wigner-Seitz cell of the reciprocal lattice is usually called the *first Brillouin zone*. In other words, it is the set of points lying closer to $\underline{K} = 0$ than to any other reciprocal lattice point. It can be constructed in exactly the same way as the Wigner-Seitz cell of the direct lattice: draw the perpendicular bisectors of all the vectors which join the central reciprocal lattice point to all others, the first Brillouin zone is the volume of reciprocal space that can be reached from the origin without crossing any of these bisecting planes.

We will start using the first BZ quite a lot before too long, and you will start seeing the high symmetry points in it given funny symbols like ?, X, Σ etc. Don't worry about this. They are meaningful in the sense that their origin lies in group theory, but we can treat them as labels without worrying about their meaning.

In reciprocal space we call these bisecting planes *Bragg planes* for reasons that will soon become apparent. The *second* BZ is that set of points that is separated from the origin by one Bragg plane. Similarly, the ith BZ is the set of points that is separated from the origin by $i-1$ Bragg planes. It is important to realise that, despite their variety of strange shapes, each BZ is a primitive cell of the reciprocal lattice. This follows from the fact that each point in reciprocal space [27] has a uniquely defined ith nearest neighbour, and so belongs to the ith BZ of precisely one lattice point. Convince yourself that this works for the 2D square lattice shown in Fig. 2.24.

It is worth pausing here to make an observation. Let's take the simple cubic Bravais lattice with lattice constant L. The first BZ is just a cube side $2\pi/L$ centred on $\underline{K} = 0$. So the minimum distance from the origin to the first BZ boundary, which we can call k_{BZ}, is π/L. In the free electron model the conduction electron density n, which is of order L^{-3}, is equal to $k_f^3/(3\pi^2)$. It follows that $k_{BZ} \sim k_f$. So what? This looks rather unpromising, but will turn out to be exceedingly important.

2.5.6 Lattice planes and Miller indices revisited

Earlier we mentioned the way *Miller indices* of planes are obtained. Here we will generate an alternative (but consistent) definition. A Bravais lattice can be divided into families of parallel equally spaced planes of lattice sites, as shown

[27] Except those points on BZ boundaries.

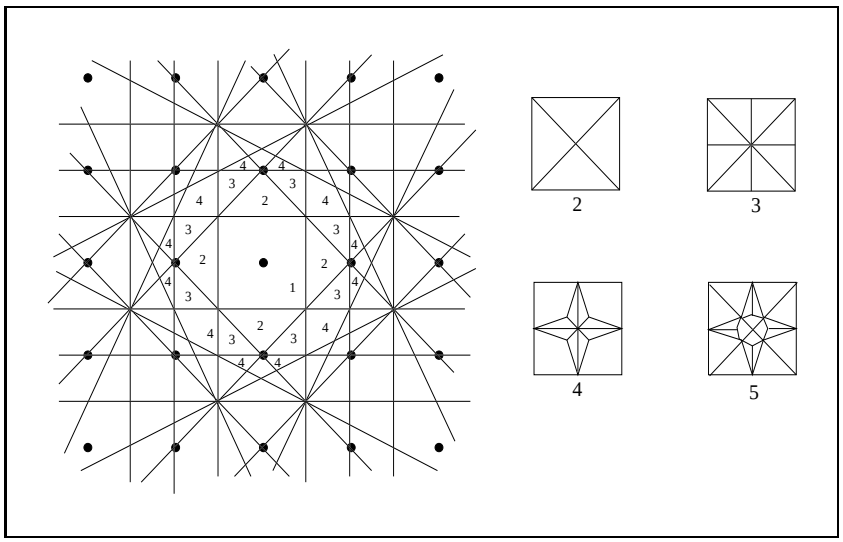

Figure 2.24: Construction of Brillouin zones for the 2D square lattice.

in Fig. 2.25 on the following page. The intimate relationship between reciprocal lattice vectors and crystal planes is embodied in the theorem:

For any family of lattice planes separated by d, there are reciprocal lattice vectors perpendicular to the planes, the shortest of which have length $2\pi/d$. Conversely, for any reciprocal lattice vector $\underline{K}$ there is a family of lattice planes normal to $\underline{K}$ and separated by a distance d, where $2\pi/d$ is the length of the shortest reciprocal lattice vector parallel to $\underline{K}$.

This rather cumbersome theorem will be useful when we consider x-ray diffraction.

To specify the orientation of a plane one usually quotes a vector normal to it. Since we know that there are reciprocal lattice vectors which are normal to any family of lattice planes, it is natural to use the smallest suitable reciprocal lattice vector to represent the normal. We could define the Miller indices of a plane to be the components (in terms of primitive reciprocal lattice vectors $\underline{a}^*, \underline{b}^*$ and $\underline{c}^*$) of the shortest $\underline{K}$ vector normal to the plane. Thus:

A plane with Miller indices $(h\,k\,l)$ is normal to the reciprocal lattice vector $\underline{K} = h\underline{a}^* + k\underline{b}^* + l\underline{c}^*$.

Is this consistent with our earlier definition? The equation of a plane normal to $\underline{K}$ is $\underline{K}.\underline{r} = A$, where A is some constant. Since the planes contain lattice points, $A = 2\pi \times$ integer, and for the plane closest to the origin (without including it) we have simply $A = 2\pi$. The intercepts of this plane on the axes given by the primitive vectors of the direct lattice we denote u, v, w (not necessarily integers), which implies

$$u\underline{K}.\underline{a} = v\underline{K}.\underline{b} = w\underline{K}.\underline{c} = 2\pi. \tag{2.26}$$

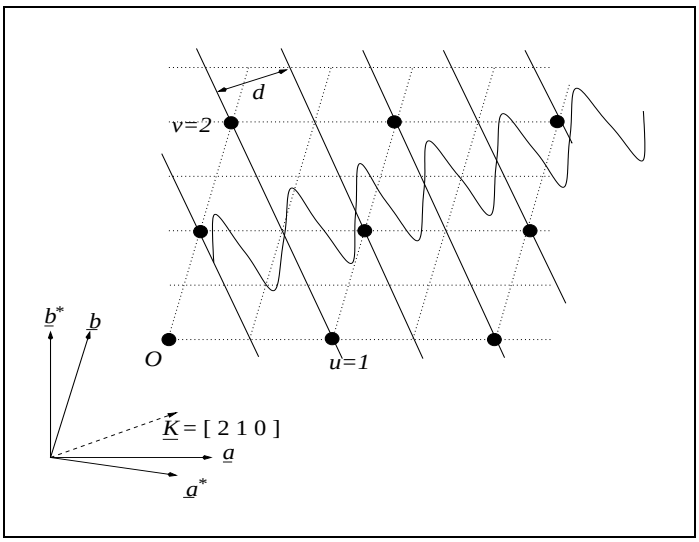

Figure 2.25: A slice through a 3D Bravais lattice (which is called "monoclinic" but don't worry about that) which we have decomposed into (2 1 0) planes, shown by the solid lines. Also shown is a plane wave with wavevector perpendicular to the planes and with magnitude equal to $2\pi/d$.

Since $\underline{K} = h\underline{a}^* + k\underline{b}^* + l\underline{c}^*$ and using Eq. 2.23 on page 48, it follows that

$$u = 1/h,\ v = 1/k,\ w = 1/l. \tag{2.27}$$

This is precisely our earlier definition.

2.6 Diffraction

So far we have discussed some of the common crystal structures and how they can be described. Most of our detailed knowledge of the structure of condensed matter has come from experiment and in this section we will see how (and Edinburgh has an unparalleled reputation in this field) quite soon.

In 1912 Max von Laue pointed out that since crystals are periodic they should act as diffraction gratings for incident radiation with wavelength comparable to interatomic distances. For electromagnetic radiation the dispersion relation $E = hc/\lambda$ suggests the use of x-rays with energy around 10 keV. For (non-relativistic) beams of massive particles the dispersion relation is $E = p^2/2m = \hbar^2 k^2/2m$, implying energies of 100 eV and 20 meV are suitable for electron and neutron diffraction respectively. X-rays and neutrons interact quite weakly with solids and are able to penetrate relatively long distances. A description in terms of single scattering events is therefore appropriate. Electrons on the other hand interact strongly with solids, penetrating only a few nm and making electron diffraction an important technique in surface physics. Here we concentrate on the diffraction of x-rays.

Before we start, we should mention a few assumptions.

• We will take a quantum mechanical view solely in the sense that x-rays (or neutrons or electrons) have a wave-like nature.

• We assume that each photon scatters off only one unit cell.[28]
• We have not identified any mechanism for the crystal to absorb any on the energy of the x-ray beam, and so the scattering must be elastic i.e. the photon energy (or wavelength) of incident and scattered x-rays is the same.

2.6.1 Bragg formulation

In 1913, W.H. and W.L. Bragg discovered that crystalline solids gave quite characteristic patterns of reflected x-rays in which strong reflected beams were observed for specific wavelengths and for specific crystal orientations. These were accounted for by assuming that
(i) x-rays are specularly reflected from lattice planes, and
(ii) successive planes give rise to constructive interference.
By considering the rays reflected from two successive lattice planes (see Fig. 2.26) one can see that the 'Bragg condition' means that the path difference for the two rays must be an integer multiple of the wavelength:

$$n\lambda = 2d\sin\theta \tag{2.28}$$

where n is called the *order* of the reflection.

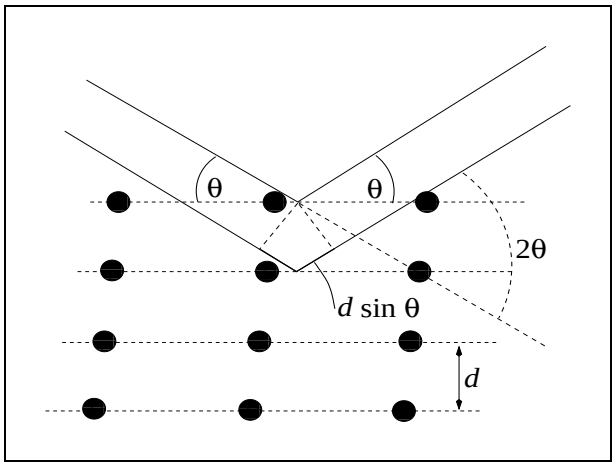

Figure 2.26: Bragg reflection from a family of lattice planes. (The dots represent lattice sites, not atoms.)

2.6.2 Von Laue formulation

In this approach we consider the scattering from each unit cell. We expect strong reflections for those wavelengths and geometries which give rise to constructive interference from all unit cells of the Bravais lattice. We start by considering just two lattice sites separated by a (real space) Bravais lattice vector $\underline{R}$, as shown in Fig. 2.27. The path difference in this case is $\underline{R}.(\hat{\underline{n}} - \hat{\underline{n}}')$, which again

[28] This is sometimes known as the "kinematic approximation". It corresponds to the first Born approximation of quantum mechanical scattering theory.

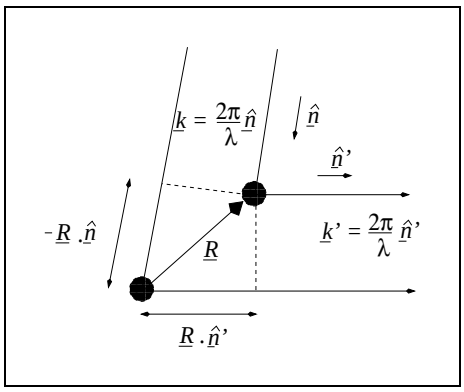

Figure 2.27: The path difference for reflection from two lattice sites separated by Bravais lattice vector $\underline{R}$.

must be an integer multiple of wavelengths for constructive interference. This leads to the Von Laue condition:

$$2\pi m = \underline{R}.(\underline{k} - \underline{k}') \tag{2.29}$$

where $\underline{k}$ and $\underline{k}'$ are the incident and outgoing wavevectors respectively, and m is an integer. If we consider the full lattice, Eq. 2.29 must apply for all Bravais lattice vectors. Comparison with Eq. 2.22 on page 48 shows that the Von Laue condition for diffraction is simply that the change in wavevector $\underline{k} - \underline{k}'$ is a reciprocal lattice vector $\underline{K}$. This being the case, we can label a diffracted beam with the h, k, l indices of the $\underline{K}$ vector relevant to that diffraction condition.

2.6.3 Bragg planes

The Von Laue scattering condition can be expressed in a more graphic way by recalling that we have assumed elastic scattering (i.e. $|\underline{k}'| = |\underline{k}|$). Equating $|\underline{k}'|^2$ with $|\underline{k}|^2$ leads to

$$\underline{k}.\hat{\underline{K}} = \frac{1}{2}K \tag{2.30}$$

where $\hat{\underline{K}}$ is a unit vector in the direction of $\underline{K}$. This means that the component of $\underline{k}$ in the direction of $\underline{K}$ is equal to half the magnitude of $\underline{K}$. In other words, the Laue condition means that the tip of the incident wavevector must lie on a plane which bisects a reciprocal lattice vector, i.e. on a *Brillouin zone boundary*, as shown in Fig. 2.28 on the facing page.

The equivalence of the Bragg and Von Laue formulations can be seen by noting that from Fig. 2.28 the magnitude of $\underline{K}$ is $2k \sin\theta$ which according to our earlier theorem is also equal to $2n\pi/d$ where d is the interplanar spacing. A Laue diffraction peak corresponding to a change in wavevector equal to reciprocal lattice vector $\underline{K}$ corresponds to a Bragg reflection from the family of direct lattice planes perpendicular to $\underline{K}$. The order of Bragg diffraction n is given by K divided by the shortest reciprocal lattice vector parallel to $\underline{K}$.

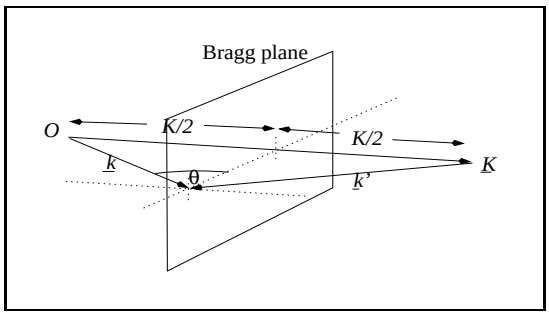

Figure 2.28: Geometric interpretation of the Von Laue diffraction condition. The tip of $\underline{k}$ must lie on a Brillouin zone boundary.

The perpendicular bisectors (i.e. the Brillouin zone boundaries) we also call *Bragg planes* as these are parallel to the lattice planes which Bragg assumed would produce specular reflection.

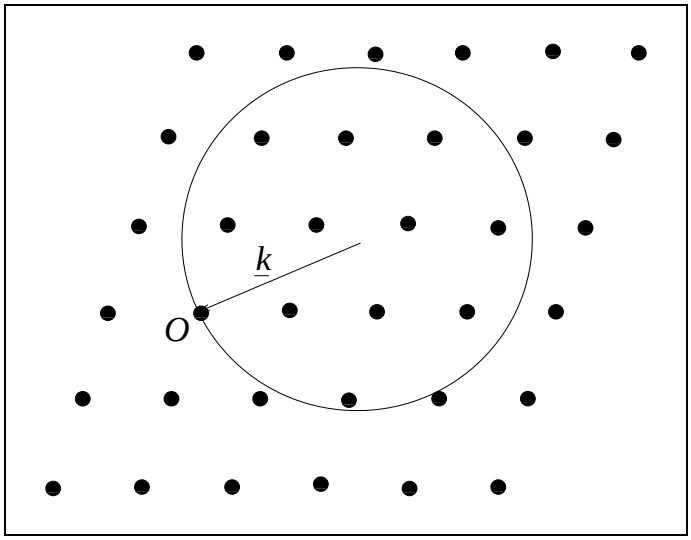

Figure 2.29: The Ewald construction. For arbitrary $\underline{k}$ and crystal orientation the diffraction conditions need not be met.

2.6.4 The Ewald construction

If we direct a monoenergetic x-ray beam onto a randomly oriented single crystal then the incident wavevector will probably not fall on a Bragg plane and so (unless we are lucky) we will not get any diffracted beams. This is easily visualised using the *Ewald construction*, shown in Fig. 2.29. First the reciprocal lattice points are drawn. We then choose a site to be the origin and draw on the incident wavevector $\underline{k}$ with its tip at the origin. After drawing a circle of radius k centred on the other end of $\underline{k}$, a little thought reveals that the Von Laue diffraction condition is satisfied if and only if the sphere hits another reciprocal lattice point (as well as the origin). In general there need be no such

points. The objective of an experimental x-ray diffraction method is to fiddle with the geometry so as to 'engineer' the diffraction condition so that structural information can be extracted.

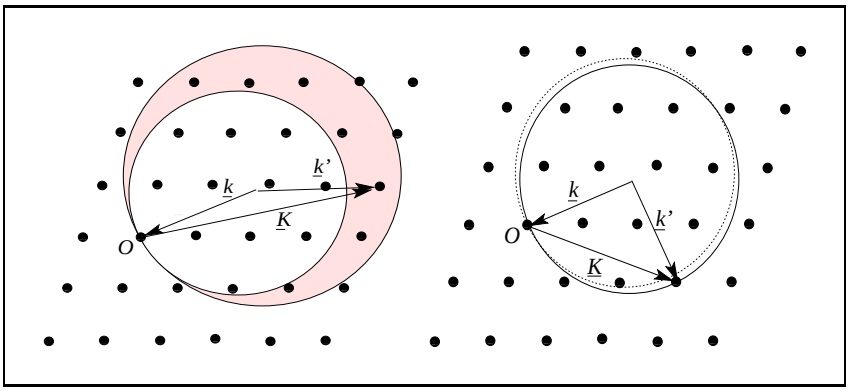

Figure 2.30: The Ewald construction illustrating the Laue (left) and the rotating crystal (right) methods.

2.6.5 Experimental x-ray diffraction

Laue method

The crystal orientation is held fixed but the incident x-ray beam contains a band of wavelengths between λ_1 and λ_2. The Ewald construction for this method shown in Fig. 2.30 consists of two spheres (each of which cuts the origin) with radii $2\pi/\lambda_2$ and $2\pi/\lambda_1$. All those k-points which lie within the outer sphere but not the smaller sphere will give rise to diffracted beams.

Rotating crystal method

Here the crystal is rotated about some axis (possibly several axes). Since the primitive vectors of the direct lattice are rotated, then so are the reciprocal lattice vectors and the reciprocal lattice itself. It is clear from Fig. 2.30 that diffracted beams will be created for particular k-points at particular angles of rotation.

Powder (Debye-Scherrer) method

In this approach the crystal must first be ground into a powder. Each *polycrystal* is large compared to the atomic scale (so diffraction occurs), but is so small that in a macroscopic sample all possible crystal orientations will be present. Diffraction rings are produced for deviation angles 2θ (see Fig. 2.26).

Whichever of these methods is used to engineer diffraction beams one must arrange for an x-ray detector (e.g. a photographic plate) to record the directions of these beams relative to the incident x-ray beam. This information allows us to determine $\underline{k}' - \underline{k}$ and so the reciprocal lattice of the sample (and hence its direct Bravais lattice). But identifying the Bravais lattice of a crystal is not

enough - we also need to know the basis if we are to determine the structure of an unknown crystal. This information is contained in the *intensities* of the diffracted beams. The diffraction condition specified in either the Bragg or von Laue form gives only the directions of the diffracted x-ray beams. To understand their intensities we need a more sophisticated approach.

2.7 Scattering Theory: crystals

The aim of this section is to understand the intensities of beams of x-rays diffracted from crystals. For now we continue to consider only perfect crystals[29] but the scattering theory introduced here will also allow treatment of amorphous solids and liquids.

At the start of Sec. 2.6 we introduced three assumptions. In this more general treatment of the scattering of waves by a crystal we will have to retain two of these, namely the "semi-classical" approach (the probe beam is wave-like but the scattering is treated classically) and the "kinematic approximation" (the probe beam is scattered only weakly so multiple scattering can be neglected).

Consider first a small volume of a crystal at position $\underline{r}$, as shown in Fig. 2.31. If a source with angular frequency ω_0 emits spherical waves from position $\underline{r}_s$ then, provided r_s is sufficiently big, plane waves with amplitude

$$A(\underline{r},t) = A_0\, e^{i[\underline{k}.(-\underline{r}_s+\underline{r})-\omega_0 t]} \tag{2.31}$$

where A_0 is a constant, will be incident on our scattering volume. This will then be a secondary source of spherical waves and the scattered amplitude reaching the detector at $\underline{r}_d$ will be

$$A'(\underline{r}_d,t) = A(\underline{r},t) \times \rho(\underline{r},t) \times \frac{e^{ik'|r_d-r|}}{|r_d-r|} \approx \frac{A_0\, e^{i(\underline{k}'.\underline{r}_d - \underline{k}.\underline{r}_s-\omega_0 t)}}{r_d}\, \rho(\underline{r},t)\, e^{i(\underline{k}-\underline{k}').\underline{r}} \tag{2.32}$$

which is clearly just the product of a spherical wave, the incident plane wave, and a scattering density[30] ρ. Now we assume[31] that the lattice is rigid and so $\rho(\underline{r},t) \to \rho(\underline{r})$. This means that the time-dependence of the scattered wave must be $\exp(-i\omega_0 t)$, and it follows that the photon energy of both incident and scattered waves is $\hbar\omega_0$. For the rigid lattice we have elastic scattering - the energy of the x-rays must be conserved since there is no mechanism by which it could be otherwise.[32]

To find the amplitude at the detector on account of scattering from the entire

[29] Note that our decomposition of a crystal into a perfectly regular Bravais lattice and a basis precludes the possibility of atomic vibrations. In effect we dealing with something like the $T = 0$ limit. But more about that in the next chapter.

[30] This is the density of whatever produces the scattering of the incident wave.

[31] In "Macromolecular Physics" next term you will see a more general treatment of scattering which shows how dynamical information can be obtained.

[32] This is not totally true. When x-rays undergo Bragg reflection their momentum perpendicular to the Bragg planes is reversed and so the crystal must recoil. A crystal is so heavy that the energy it takes away by recoiling is extremely small compared to the photon energy.

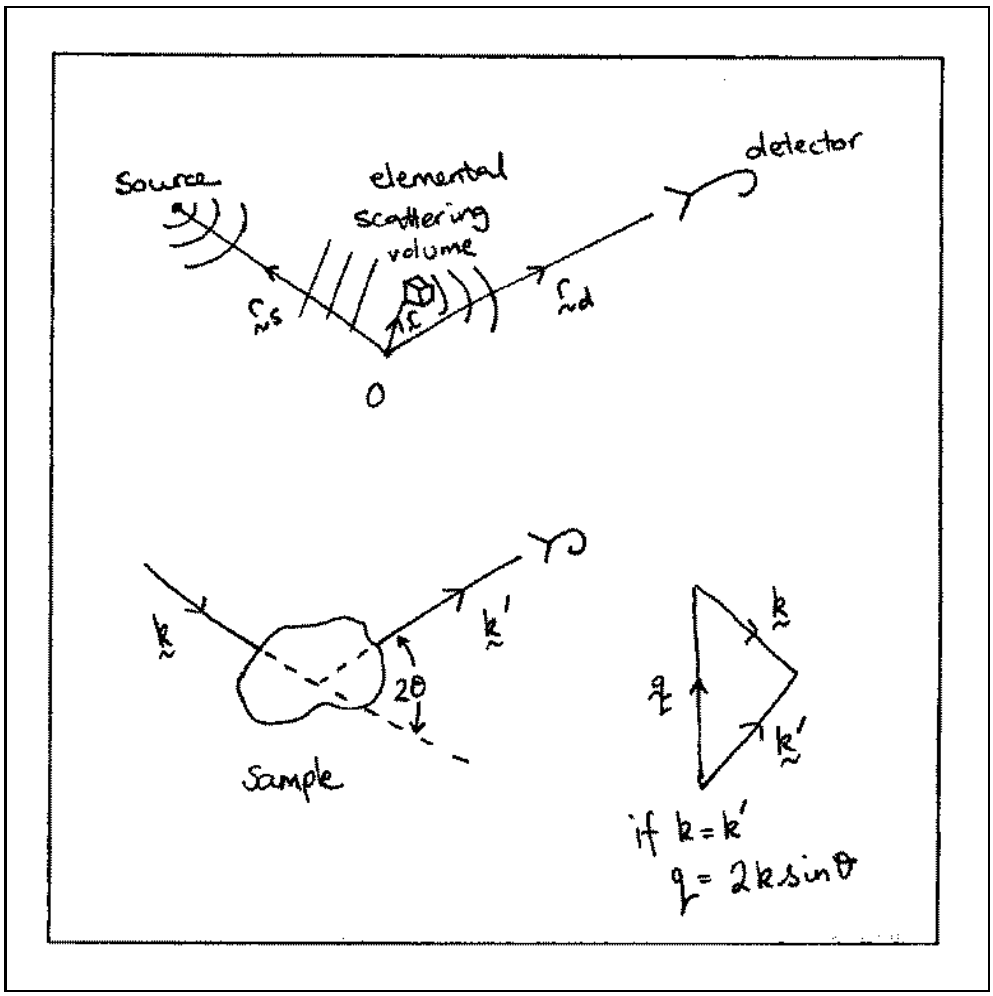

Figure 2.31: Scattering of waves from a crystal. θ is the Bragg angle.

crystal we must integrate over its volume,[33] whereby we obtain

$$A'(\underset{\sim}{Q}) \propto \int_{\text{sample}} \rho(\underset{\sim}{r})\, e^{-i\underset{\sim}{Q}.\underset{\sim}{r}}\, d\underset{\sim}{r}. \tag{2.33}$$

Note that we are now labelling the detector position by $\underset{\sim}{Q} = \underset{\sim}{k}' - \underset{\sim}{k}$, known as the *scattering vector* (see Fig. 2.31). According to Eq. 2.33,

The amplitude of the scattered wave reaching the detector is (proportional to) the Fourier Transform of the scattering density,

and this brings a number of thoughts to mind:

[33] Remember that the kinematic approximation means that we don't need to account for the scattered waves in one part of the sample getting scattered again in another region. If the probability of scattering is low then that of double scattering can reasonably be neglected.

1. It appears that after measuring A', a simple inverse Fourier transform will give us the distribution of scattering density. Unfortunately one can only measure the *intensity* of x-rays reaching the detector (which is $|A'|^2$), and so one can't " invert". Instead we must adopt the "indirect" process of guessing $\rho(\underline{r})$, calculating $|A'|^2$, and then comparing with the experimental results. Notice that two structures may give rise to the same amplitude for a particular diffracted beam, so to distinguish between them we must analyse all the beams that can be measured.

2. Eq. 2.33 has the form of Fermi's golden rule: Amplitude $\sim \langle out|\hat{O}|in\rangle$.

3. We can rewrite Eq. 2.33 as

$$A'(\underline{Q}) \propto \int_{\text{sample}} \rho(\underline{r})\, e^{-i\underline{Q}.\underline{r}}\, d\underline{r} = \left[\int_{\text{unit cell}} \rho(\underline{r})\, e^{-i\underline{Q}.\underline{r}}\, d\underline{r}\right] \times \left[\sum_{\underline{R}\in \text{BL}} e^{-i\underline{Q}.\underline{R}}\right]$$

$$= \left[\int_{\text{unit cell}} \rho(\underline{r})\, e^{-i\underline{Q}.\underline{r}}\, d\underline{r}\right] \times \left[\int_{\text{sample}} \sum_{\underline{R}\in \text{BL}} \delta(\underline{r}-\underline{R})\, e^{-i\underline{Q}.\underline{R}}\, d\underline{r}\right]. \qquad (2.34)$$

This looks like the F.T. of the contents of a unit cell (i.e. the basis) multiplied by the F.T. of the Bravais lattice vectors. This is an example of the *convolution theorem* you met last year.

4. The scattering density of a crystal is periodic and so can be expressed as a Fourier series:

$$\rho(\underline{r}) = \sum_{\underline{K}} \rho_{\underline{K}}\, e^{i\underline{K}.\underline{r}}. \qquad (2.35)$$

Inserting this in Eq. 2.33 yields

$$A'(\underline{Q}) \propto \sum_{\underline{K}} \int_{\text{sample}} \rho_{\underline{K}}\, e^{i(\underline{K}-\underline{Q})\cdot\underline{r}}\, d\underline{r} = \sum_{\underline{K}} \rho_{\underline{K}} \left\{\int_{\text{sample}} e^{i(\underline{K}-\underline{Q})\cdot\underline{r}}\, d\underline{r}\right\}. \qquad (2.36)$$

You may recognise the integral in the curly brackets as the delta function $\delta(\underline{K}-\underline{Q})$, which means that scattered amplitude only reaches the detector if the scattering vector equals a reciprocal lattice vector.[34]

We have now derived two important results:
(i) the von Laue diffraction condition:

$$\underline{Q} = \underline{K}, \qquad (2.37)$$

(ii) the amplitude of each diffracted beam is just the corresponding Fourier component of the scattering density

$$A'(\underline{K}) \propto \rho_{\underline{K}}. \qquad (2.38)$$

[34] To be precise, this is only the delta function if the integral is over all space. Since real specimens are finite, the diffracted beams are produced for a very small but non-zero range of $\underline{Q}$.

Notice that we have all these concepts in the diffraction of light from a diffraction grating. The Fourier components of the aperture play the role of $\rho_{\underline{K}}$, the total width of the grating determines the width of the diffracted beams, and the spacing between the diffracted beams is determined by precisely the same diffraction condition.

2.7.1 The structure factor and the atomic form factor

In the diffraction world $\rho_{\underline{K}}$ is often referred to as the *structure factor*, but it is nothing other than the inverse FT of $\rho(\underline{r})$:

$$\rho_{\underline{K}} = \frac{1}{V} \int_{\text{unit cell}} \rho(\underline{r})\, e^{-i\underline{K}.\underline{r}}\, d\underline{r}. \tag{2.39}$$

We can divide the scattering density into contributions from each atom within the unit cell:

$$\rho_{\underline{K}} = \frac{1}{V} \sum_{\alpha} e^{-i\underline{K}.\underline{r}_\alpha} \left\{ \int_{\text{atomic radius}} \rho_\alpha(\underline{r})\, e^{-i\underline{K}.\underline{r}}\, d\underline{r} \right\} = \frac{1}{V} \sum_{\alpha} e^{-i\underline{K}.\underline{r}_\alpha} b_\alpha(\underline{K}). \tag{2.40}$$

The term in curly brackets describes the scattering from atom alpha in the unit cell and it is usually called the *atomic form factor* and is denoted $b(\underline{Q})$ for scattering vector $\underline{Q}$.

The time has come to be more specific about the nature of the scattering density. X-rays are scattered by the charges in a solid. Standard scattering theory shows that the strength of this interaction scales inversely with the mass of the charged particle involved. Thus we need only consider x-ray scattering by the electron distribution (i.e. $\rho(\underline{r})$ is simply the distribution of electrons) and two important observations follow. Firstly, x-ray diffraction is much better at locating atoms with high atomic number. For example, hydrogen is a factor of 36 harder to see than carbon. Secondly, since only the very outermost electrons participate in chemical bonding the atomic form factor is approximately independent of chemical environment. This means the form factor can be computed for each atom from a Hartree-Fock calculation.[35]

Since atoms are spherical, the atomic form factor is given by

$$b(Q) = \int \rho(r)\, 4\pi r^2\, \frac{\sin(Qr)}{Qr}\, dr. \tag{2.41}$$

For $Q = 0$ (i.e. $\theta = 0$, see Fig. 2.31), b is equal to the atomic number Z. But as the length of the scattering vector $\underline{Q}$ increases (or equivalently, as $k \sin\theta$ increases), $b(Q)$ falls away, as shown in Fig. 2.32. We can readily understand this figure. We know that the electron density of an atom is roughly 2 Å wide and so we expect its Fourier transform to be $2\pi/2$ Å^{-1} wide in Q space. This is what we see. The fading away of diffraction intensities at high Q can be a bit frustrating as it makes it hard to measure some of the diffracted beams.

Neutrons are scattered by the nuclei of the atoms in a solid. These are extremely small on the scale of the neutron wavelength in typical neutron diffraction experiments and so the atomic form factor for neutron scattering is essentially independent of Q (the FT of a delta function is a constant). Another

[35] This is just a method for calculating the wavefunctions of the electrons in atoms, as discussed in "Atomic and Molecular Physics".

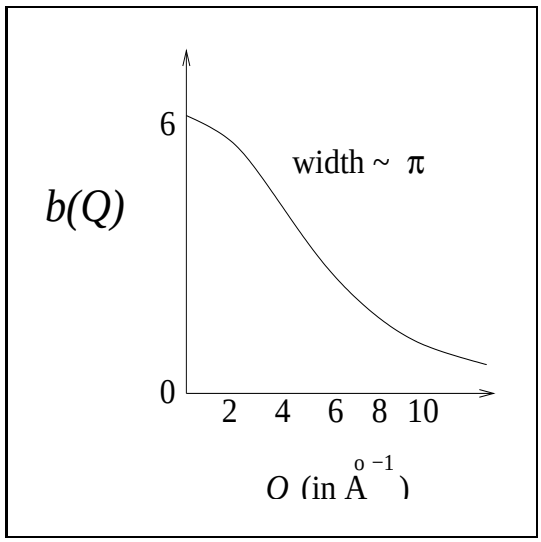

Figure 2.32: The atomic form factor of carbon for x-ray scattering.

advantage of neutron diffraction is that the form factors of most elements are similar in magnitude. We can't offer a simple explanation for this, but it means that neutrons are good at finding light atoms. We will see yet another appealing aspect of neutron diffraction in the next chapter.

2.7.2 Absent reflections

We noted in Sec. 2.4.3 that when expressing a crystal structure as a Bravais lattice and a basis their is no unique choice. If we choose two different Bravais lattices (each with an appropriate basis) for a given crystal structure, these choices will lead to different reciprocal lattices. The diffraction condition states that diffracted beams are produced when the scattering vector $\underline{Q}$ equals a reciprocal lattice vector. But a given crystal structure will give rise to a unique diffraction pattern. How do we reconcile these facts? The answer lies in the structure factor.

Let's consider the simple cubic crystal structure. Suppose we decide to use the most obvious choice - a simple cubic Bravais lattice with a separation of L in each direction. We can choose the set $L\underline{\hat{x}}, L\underline{\hat{y}}, L\underline{\hat{z}}$ to be our primitive vectors, and we can use a basis consisting of one atom at the origin of the unit cell.
The reciprocal of our Bravais lattice is a simple cubic lattice with sides $2\pi/L$, and so the reciprocal lattice vectors have the form $\underline{K} = h\underline{a}^* + k\underline{b}^* + l\underline{c}^*$ where h, k, l are integers and the primitive reciprocal lattice vectors are $\underline{a}^* = 2\pi/L\underline{\hat{x}}, \underline{b}^* = 2\pi/L\underline{\hat{y}}, \underline{c}^* = 2\pi/L\underline{\hat{z}}$.
We get diffraction whenever the change in wavevector $\underline{k} - \underline{k}'$ is equal to a reciprocal lattice vector $\underline{K}$. Since we have only one atom per unit cell, the structure factor is the same for all $\underline{K}$.

Now let's take a unit cell with dimension $2L$. We can choose the set $2L\underline{\hat{x}}, 2L\underline{\hat{y}}, 2L\underline{\hat{z}}$ to be our primitive vectors. The Bravais lattice is again simple cubic, but now the basis is consists of eight atoms at coordinates:
$(0,0,0), (L,0,0), (0,L,0), (L,L,0), (0,0,L), (L,0,L), (0,L,L), (L,L,L)$.

The reciprocal of the Bravais lattice is now a simple cubic lattice with sides $2\pi/2L = \pi/L$. (Remember that a larger direct unit cell gives a smaller reciprocal lattice unit cell.) Primitive vectors for the new reciprocal lattice are $\underline{a}^* = \pi/L\underline{\hat{x}}, \underline{b}^* = \pi/L\underline{\hat{y}}, \underline{c}^* = \pi/L\underline{\hat{z}}$.
We get diffraction whenever the change in wavevector $\underline{k} - \underline{k}'$ is equal to a reciprocal lattice vector $\underline{K}$. This is a bit worrying because we now have 8 times as many reciprocal lattice vectors as we had before. Does this mean we should get 8 times as many diffraction beams?

Let's finish the job and calculate the structure factor. Using Eq. 2.39 we get

$$\frac{\rho_{\underline{K}}}{b(\underline{K})} = \exp(0) + \exp[-i\pi(100).(hkl)] + \exp[-i\pi(010).(hkl)] + \exp[-i\pi(110).(hkl)]$$

$$+ \exp[-i\pi(001).(hkl)] + \exp[-i\pi(101).(hkl)] + \exp[-i\pi(011).(hkl)] + \exp[-i\pi(111).(hkl)]$$

$$= 1 + \exp[-i\pi h] + \exp[-i\pi k] + \exp[-i\pi(h+k)] + \exp[-i\pi l] + \exp[-i\pi(h+l)]$$

$$+ \exp[-i\pi(k+l)] + \exp[-i\pi(h+k+l) \tag{2.42}$$

where b is the atomic form factor and a shorthand notation has been used for the vectors in the exponentials. Unless h, k and l are all even, $\rho_{\underline{K}}$ equals zero. This means that all the extra diffracted beams have zero intensity.

Although it is a bit worrying that we have choice in which unit cell to use (in fact there is an infinite number of alternative unit cells), the diffraction properties don't depend on the choice.

Another example: the BCC crystal structure

We can treat the BCC crystal structure as a simple cubic Bravais lattice with a two atom basis. To be more specific we have the Bravais lattice vectors $\underline{a} = L\underline{\hat{x}}, \underline{b} = L\underline{\hat{y}}, \underline{c} = L\underline{\hat{z}}$ and a basis consisting of atoms at $\underline{d}_1 = \underline{0}, \underline{d}_2 = (\underline{\hat{x}}+\underline{\hat{y}}+\underline{\hat{z}})L/2$. We know that the reciprocal lattice is simple cubic with side length $2\pi/L$ but we must now evaluate the structure factor associated with each Bragg reflection. In this case,

$$\frac{\rho_{\underline{K}}}{b(\underline{K})} = e^{-i\underline{K}.\underline{0}} + e^{-iL\underline{K}.(\underline{\hat{x}}+\underline{\hat{y}}+\underline{\hat{z}})/2}. \tag{2.43}$$

Since all the reciprocal lattice vectors of a simple cubic Bravais lattice have the form

$$\underline{K} = \frac{2\pi}{L}\left(h\underline{\hat{x}} + k\underline{\hat{y}} + l\underline{\hat{z}}\right) \tag{2.44}$$

it follows that

$$\rho_{\underline{K}} = f\left(1 + e^{-i\pi(h+k+l)}\right). \tag{2.45}$$

$\rho_{\underline{K}}$ equals 2f if $h+k+l$ is an even number, and equals 0 if $h+k+l$ is odd, and so half of the expected diffraction beams will be absent.

To understand why this should be, recall that the condition for diffraction is that the change in wavevector must be equal to a reciprocal lattice vector.

The reciprocal of the simple cubic Bravais lattice with sides of length L is itself a simple cubic lattice but with lattice parameter $2\pi/L$. The reciprocal lattice points with non-zero structure factor form an FCC lattice with lattice parameter $4\pi/L$. This is precisely the reciprocal lattice we would have obtained by treated the BCC crystal structure as a BCC Bravais lattice with a single lattice site per unit cell.

2.7.3 Almost absent reflections

(These are explored in detail in Problem Sheet 2.)
The effect of the atomic form factor can be seen by considering diffraction from CsCl. This compound is simple cubic with a two atom basis. The only difference between this and the BCC structure is that the two atoms in the basis are of different elements. Although the exponentials in the structure factor will tend to cancel beams when $h+k+l$ is odd (as for the BCC structure), the cancellation is not perfect since Cs and Cl have different atomic form factors. However, if we had chosen CsI then the cancellation is almost perfect since Cs^+ and I^- are iso-electronic and so have very similar atomic form factors.

In problems like this remember that the *intensity* of dffracted beams is what is actually measured, not the amplitude.

2.7.4 Accidental absences

The absences discussed in Sec. 2.7.2 arose when we decided to deal with non-primitive unit cells. These *systematic absences* are the result of destructive interference from *identical* atoms in each non-primitive unit cell.

Accidental absences in the diffraction patterns of crystals can occur if the primitive unit cell contains more than one atom *of the same type*. But notice that even atoms of the same type in a multi-atom primitive unit cell are not (by definition) crystallographically equivalent. If they were then the unit cell would not be primitive. On account of this crystallographic inequivalence, the atomic form factors of the atoms in the primitive unit cell are not rigorously identical, but, since the atoms are of the same type, very nearly so. As a result certain diffraction beams have negligible intensity, but if the atomic form factors are sufficiently different some accidental "absences" do show up as very weak beams.

2.8 Scattering theory: non-crystalline matter

In the previous section we saw that diffraction is the result of constructive interference of waves scattered from each unit cell in a crystal. In fact we only considered the extreme case of perfect infinite crystals with strictly rigid atomic positions. One might guess that allowing the atoms to jiggle around a bit will lead to a slight spoiling of the diffraction patterns expected for rigid crystals. This is basically correct, but we will see an additional feature in the next Chapter. In this section we will try to understand if scattering experiments can provide any information on the structure of non-crystalline matter.

2.8.1 "Everyday" scattering

When a wave travels through a homogeneous medium nothing very interesting happens - it just keeps going with speed determined by the properties of the material (the refractive index in the case of electromagnetic waves). In the Huygens approach to optics this is just scattered "wavelets" interfering constructively in the forward direction. When a new material is encountered the interference of the wavelets gives rise to reflected and refracted rays at the interface. So although water is clear we can still detect its presence in a cup or glass due to the "meniscus" at the water-air interface. This idea that scattering is produced by *variations* in refractive index is quite general.

Of course all matter exhibits variations when one looks on a sufficiently fine scale; the important question is whether the material is homogeneous *on the scale of the wavelength* of the incident wave. Thus water (or any solution) appears homogeneous to a light wave. Milk on the other hand contains droplets of fat which are comparable in radius to λ for visible light (in fact a bit larger). As far as light waves are concerned milk is not homogeneous and since the fat droplets are not ordered light scattering gives rise to a characteristic "milky" appearance. Similarly fog appears inpenetrable and clouds appear white.

In the nineteenth century Tyndall found that for the special case of a medium which is inhomogeneous on the scale of λ but when the particle size is small on this scale the scattered light tends to be slightly bluish while the transmitted beam is slightly red. Rayleigh did the mathematics and found that the scattering rate in the small particle regime varies as λ^{-4}. Rayleigh scattering (sometimes called the "Tyndall effect") explains why light scattered by molecules in the rarefied upper atmosphere produces blue skies and red sunsets, why babies' eyes are blue at birth, and why cigarette smoke has a bluish tinge.

As well as revealing the length scale of density variations, a detailed analysis of scattering can also yield quantitative information on particle *distributions*. For colloidal systems particle sizes are around the micron scale and so visible light scattering is rather informative, as exemplified by the Soft Condensed Matter research group, and Dr Egelhaaf's lectures next term. But to get structural information at the atomic scale we must resort to x-rays again.

2.8.2 Clusters and polycrystals

Consider now x-ray scattering from a finite sized crystal. The integral on the far right of Eq. 2.36 on page 59 is now no longer a strict delta function but is broadened a bit. The broadening scales inversely with the crystal size and, as we mentioned at the bottom of page 59, this effect has a precise anologue in ordinary ruled diffraction gratings of classical optics.

A polycrystalline sample or a powder could be considered to be a very large collection of clusters. The diffraction pattern from such a sample would then be the patterns of each crystal superimposed on each other. Instead of diffracted beams in particular directions we would expect continuous diffraction rings as described in Sec. 2.6.5 under "Debye-Scherrer". Now the size of the cluster will determine the sharpness of the rings: the smaller the clusters the more blurred will be the rings, and one might then wonder what would happen if the clusters were exceedingly small, say just a few atoms across. In fact this is a reasonable model of a liquid. Liquids are not periodic so they shouldn't give diffraction

spots, but they are not totally devoid of structure, as we observed in Sec. 2.3.1 on page 30. Can we learn anything useful by scattering x-rays off liquids?

2.8.3 amorphous solids and liquids

For notational simplicity we will consider in this section non-crystalline matter comprising N identical atoms, and we'll imagine freezing the atoms in place. The "dynamic" version can be seen in "Macromolecular Physics" next term. We start be rewriting the expression for the scattering amplitude in the first line of Eq. 2.34 on page 59 in the form

$$A'(\underline{Q}) = \left[\int_{\text{atom}} \rho(\underline{r})\, e^{-i\underline{Q}\cdot\underline{r}}\, d\underline{r}\right] \times \left[\sum_{i=1}^{N} e^{-i\underline{Q}\cdot\underline{r}_i}\right] \tag{2.46}$$

where the index i labels the atoms and $\underline{r}_i$ are the atomic coordinates.[36] The scattered *intensity* is then $I(\underline{Q}) = |A'(\underline{Q})|^2$. Using our earlier definition of the atomic form factor $f(\underline{Q})$ we can write

$$I(\underline{Q}) \propto |b(\underline{Q})|^2 \times \left|\sum_i e^{-i\underline{Q}\cdot\underline{r}_i}\right|^2 = |b(\underline{Q})|^2 \times \left[\sum_i e^{-i\underline{Q}\cdot\underline{r}_i}\right] \times \left[\sum_j e^{+i\underline{Q}\cdot\underline{r}_j}\right]$$

$$= N|b(\underline{Q})|^2 \times \left[\frac{1}{N}\sum_{i,j} e^{-i\underline{Q}\cdot(\underline{r}_i - \underline{r}_j)}\right]. \tag{2.47}$$

The term in brackets on the far right hand side is usually called the "liquid structure factor",[37] denoted $S(\underline{Q})$. Using the integral form of the delta function we could express S as

$$S(\underline{Q}) = \sum_{i,j} \int e^{-i\underline{Q}\cdot\underline{r}} \left(\frac{1}{N}\delta(\underline{r} - [\underline{r}_i - \underline{r}_j])\right)\, d\underline{r}$$

$$= 1 + \sum_{i\neq j} \int e^{-i\underline{Q}\cdot\underline{r}} \left(\frac{1}{N}\delta(\underline{r} - [\underline{r}_i - \underline{r}_j])\right)\, d\underline{r}. \tag{2.48}$$

The sum of delta functions here looks reminiscent of the definition of the radial distribution function (Eq. 2.13 on page 32), and we can make this explicit by further manipulation:

$$S(\underline{Q}) = 1 + \sum_{i\neq j} \int e^{-i\underline{Q}\cdot\underline{r}} \left(\frac{1}{N}\delta(\underline{r} - [\underline{r}_i - \underline{r}_j]) - 1\right)\, d\underline{r} + \int e^{-i\underline{Q}\cdot\underline{r}}\, d\underline{r}$$

$$= 1 + \int [g(\underline{r}) - 1]\, e^{-i\underline{Q}\cdot\underline{r}}\, d\underline{r} + \delta(\underline{Q}). \tag{2.49}$$

[36] Notice that we reserve the symbol $\underline{R}$ for the Bravais lattice vectors.

[37] Notice that the "structure factor" for a crystal, earlier denoted by $\rho_{\underline{Q}}$, is a scattering *amplitude* while the liquid structure factor $S(\underline{Q})$ is an *intensity*. Unfortunately the research literature for liquids and crystals use inconsistent terminology.

The delta function means that we see a strong x-ray intensity when the scattering vector $\underline{Q}$ is zero. This is the unscattered beam arising from our assumption of weak scattering. The important message in the above expression is that the liquid structure factor, and hence the scattering intensity, measure the Fourier transform of $g(\underline{r})-1$, i.e. the departure from homogeneity.[38] Taking the Fourier inverse gives the RDF.

The experimental liquid structure factor for liquid argon is shown in Figure 2.33 and can be seen to be dominated by a single peak. Thus the scattering pattern of a typical liquid is one reasonably well defined ring. S at small $\underline{Q}$ is a measure of the long wavelength structure in the RDF: for a liquid there is none and so S is negligible. The peak in S occurs when $2\pi/Q$ equals the nearest neighbour distance.

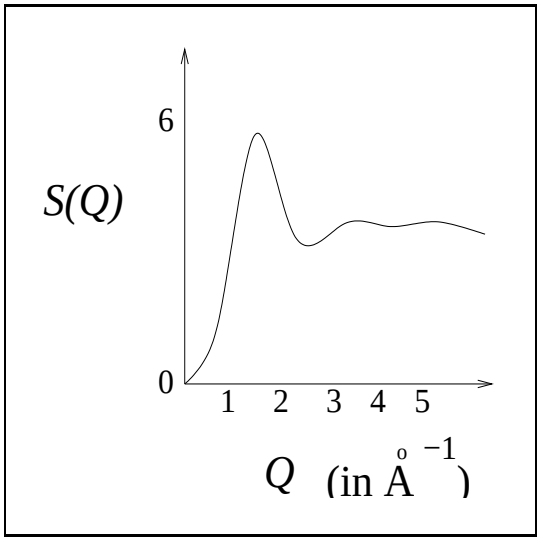

Figure 2.33: The liquid structure factor of liquid argon.

[38] Recall that $g = 1$ for a homogeneous fluid.

Chapter 3

Crystal dynamics

We started the course with the free electron gas model of the metallic state, the classical version of which led us into some trouble. It predicted a contribution of some $\frac{3}{2}nk_b$, where n is the electron density, to the heat capacity of metals and a zero contribution to non-metals. At room temperature the heat capacities of metals and non-metals alike are of order $3Nk_b$, where N is the number of atoms per unit volume. Clearly the electronic contribution is much smaller than predicted and there is a large contribution which we have not explained. We figured out the problem with the electronic contribution (with Fermi-Dirac statistics the predicted contribution is much smaller), and we saw some experimental data that supported our revised theory. This data also showed that the second contribution to the C_V of metals (the sole contribution in non-metals) varies as T^3 at low temperature. We need to understand the origin of this effect since it appears to be the dominant mechanism for the thermal excitation of solids under most circumstances.

On the subject of thermal properties, we have yet to understand how non-metals conduct heat. It is clear that the conduction electrons do most of the job in metals but non-metals *do* conduct heat to some extent. How?

Our understanding of electrical resistance also suffered a knock in Chapter 1. We started by attributing the scattering of electrical currents to collisions between the electrons and the much larger ions. After accounting for FD statistics we deduced a mean distance between collisions which was a factor of 25 or so too big for this mechanism to make sense. So what scatters the conduction electrons in metals?

To establish a basis for a more sophisticated treatment of electrons in solids, and in particular the electron-ion interactions, we spent some time in Chapter 2 getting to grips with the structure of condensed matter. We will return to electronic structure in the next chapter where we hope that some explanation of the existence of metals, insulators, semiconductors and superconductors will emerge. As well as using our understanding and knowledge of crystal structures in attempting to describe the behaviour of electrons in condensed matter, one might also hope to understand why particular elements adopt particular structures in the first place. This is a job for "Solid state physics", but thinking along these lines will enable us to advance the story here as well.

We saw in the last chapter that the potential energy of two noble gas atoms as a function of their separation is reasonably well-described by the Lennard-

Jones equation. Taking the minimum of the L-J well we found that we could get a good understanding of the bond lengths and cohesive energies of the noble solids. So much for the ground state. What happens when $T > 0$? Of course we know that the atoms are no longer confined to the bottom of the L-J well, but rather they jiggle around. The greater the T, the more jiggling goes on.

The Lennard-Jones potential is of rather limited use, but one may still expect that the potential energy of any crystal as a function of the interatomic distance would look qualitatively similar, i.e. like a well centred on the ground state. In this chapter we will not be too concerned about the detailed form of this well. We will find that many of the properties of vibrating crystals can be understood without such detailed knowledge. In fact we will explore the main concepts using a simple model of atoms connected together to nearest neighbours by springs. It is tempting to associate these springs with chemical bonds, but such a physical interpretation is not always appropriate. Nor is it necessary since more complicated theories which account for longer range interactions lead to substantially similar results.

The notion of vibrations in crystals seems like a very promising avenue for us now. These vibrations should be able to store energy and so should be relevant to the heat capacity problem. Furthermore, if we displace an atom then we would expect the disturbance to spread through the crystal like a wave. In other words, we have a mechanism for thermal conduction. The thought that crystals are not rigid requires no great leap of the imagination. After all it accounts for the fact that crystals have a finite compressibility, rigidity and shear strength. In effect we are exploring the origins of these macroscopic phenomena. But the subtle interplay between crystal vibrations and other types of waves (electromagnetic, neutrons and electrons) will give us some new physics as well.

Before we start the serious business, one more observation is appropriate: condensed matter conducts sound waves. These waves have wavelengths which are enormous compared to the size of atoms and they are usually treated using macroscopic elastic theory (Hooke's law etc.). This shows that the displacement u induced by longitudinal sound waves satisfies an equation of the form

$$\frac{C}{\rho}\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 u}{\partial t^2} \tag{3.1}$$

where x is distance in the direction of the wave, C is the modulus of elasticity (something like Young's modulus) and ρ is the density of the crystal, and so the speed of sound is $\sqrt{C/\rho}$. This is typically of order of a few thousand ms^{-1}.

We might also note here that for plane waves $\exp(ikx - \omega t)$ the wave velocity (sometimes called the *phase velocity*) is ω/k. For sound waves this is independent of wavelength. Whatever λ (or k), the wave velocity is the same. This is rather like the behaviour of light in vacuum, but different to what we saw for free electron waves, for which $\omega = \epsilon/\hbar = \hbar k^2/(2m)$. In general we call the relation between ω and k a *dispersion relation*. For dispersive waves (for which the phase velocity depends on k) the energy of a superposition of component waves is transported at the *group velocity*, given by $d\omega/dk$. Obviously the group and phase velocities are equal for a non-dispersive wave.

3.1 The many-body problem, coupled oscillators and "normal modes"

The simplified view of a solid as an array of balls connected by springs is not so simple mathematically since to understand its vibrations we need to describe the simultaneous movements of all its particles. They are interacting and so this is a many-body problem. A good strategy in many-body physics is to try to re-arrange the coupled equations of motion for the N interacting particles to give a set of independent equations. This "canonical transformation" gives us equations of motion for N fictitious non-interacting particles which we can solve one at a time. In the language of many-body physics these fictitious entities are called "elementary excitations".[1]

In reality, the best one can usually do is to find fictitious entities that are *weakly* interacting. This at least allows approximate methods such as perturbation theory to be tried. But you may already be aware that our ball and spring model is a special case, provided we can assume all the forces acting between atoms are *harmonic*, i.e. the forces are proportional to the atomic displacements. Mathematically, this means that we truncate the Taylor expansion for the potential energy of the crystal after the second order derivatives. The potential energy of two atoms at separation r becomes

$$V(r) \approx V(a) + (r-a)\left(\frac{dV}{dr}\right)_{r=a} + \frac{(r-a)^2}{2}\left(\frac{d^2V}{dr^2}\right)_{r=a} \tag{3.2}$$

where a is the interatomic separation in the ground state. It is easy to see that the linear term also disappears since $-(dV/dr)_{r=a}$ is the force on the atoms at the ground state which must be zero. (The potential is always flat at the bottom on an energy well.) So we readily obtain

$$F = -\frac{dV}{dr} = -(r-a)\left(\frac{d^2V}{dr^2}\right)_{r=a} \tag{3.3}$$

which is the well known equation for "simple harmonic motion". The second derivative of the interatomic potential energy near the bottom of the well gives us the "spring constant".

We have shown that, provided the displacement from a is small, Taylor's theorem always allows us to approximate interatomic forces by harmonic springs. This is reassuring, but the best part of it is that when we couple together lots of balls using harmonic springs it is *always* possible to rearrange the *coupled* equations of motion of the balls to give a set of *uncoupled* equations. These describe the *normal modes* of the system of coupled oscillators. If you have never been through a course on coupled oscillations don't worry - it is enough to appreciate that:
(i) in the harmonic approximation the equations of motion can always be decoupled so as to describe independent modes, and
(ii) since the normal modes don't interact, the excitation energy of a system of coupled oscillators is just the sum of the energies of the normal modes.

[1] If they resemble particles, they are often called *quasiparticles*, while wavelike entities are called *collective oscillations*.

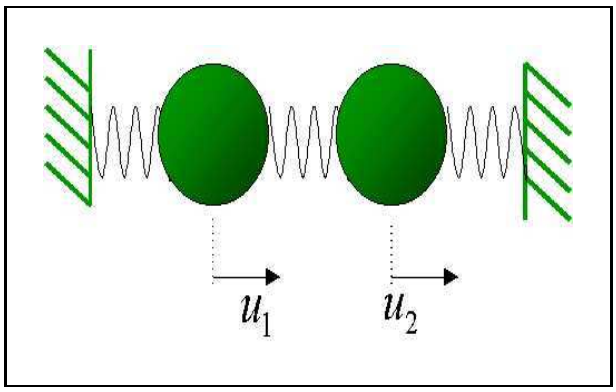

Figure 3.1: Two masses connected between fixed walls by harmonic springs.

Before considering vibrations in solids we can make a simple demonstration of the general idea of normal modes. Consider the situation shown in Fig. 3.1 of two bodies of mass m connected to each other and two stationary walls by means of three harmonic springs with spring constant S (i.e. the force exerted by each spring is $-S$ times its extension). Newton's second law gives us two equations of motion:

$$m\ddot{u}_1 = S(u_2 - u_1) - Su_1 \tag{3.4}$$

$$m\ddot{u}_2 = -Su_2 + S(u_1 - u_2) \tag{3.5}$$

where u_1 and u_2 are the displacements of the particles from equilibrium. The problem with these equations is that they each contain the displacements of *both* particles as well as a derivative: we say the equations of motion of the particles are *coupled*. If the middle spring were cut then each particle would independently execute simple harmonic vibrations when displaced from equilibrium. This does not in general happen for the coupled system but one might wonder whether it could happen in certain special cases. For this to happen it is self evident that the two particles would have to vibrate at the same frequency and it also seems likely that their vibration amplitudes and phases would have a special relationship. Let's suppose that the particle displacements during this special "mode" of vibration have the form

$$u_1 = A_1 e^{-i\omega t}, \qquad u_2 = A_2 e^{-i\omega t} \tag{3.6}$$

where A_1 and A_2 can be complex. Substituting into the equations of motion we find that ω must take either of the two values:

$$\omega_a = \sqrt{\frac{S}{m}}, \qquad \omega_b = \sqrt{\frac{3S}{m}}. \tag{3.7}$$

For the a mode we find $A_1 = A_2$; the particles move in phase with equal amplitude, while for the b mode $A_1 = -A_2$ and the particles move in antiphase.

In these two modes of vibration the balls execute SHM at the same frequency (the natural frequencies of the system) and we call them *normal modes*. It is important to appreciate that these are collective oscillations of the coupled

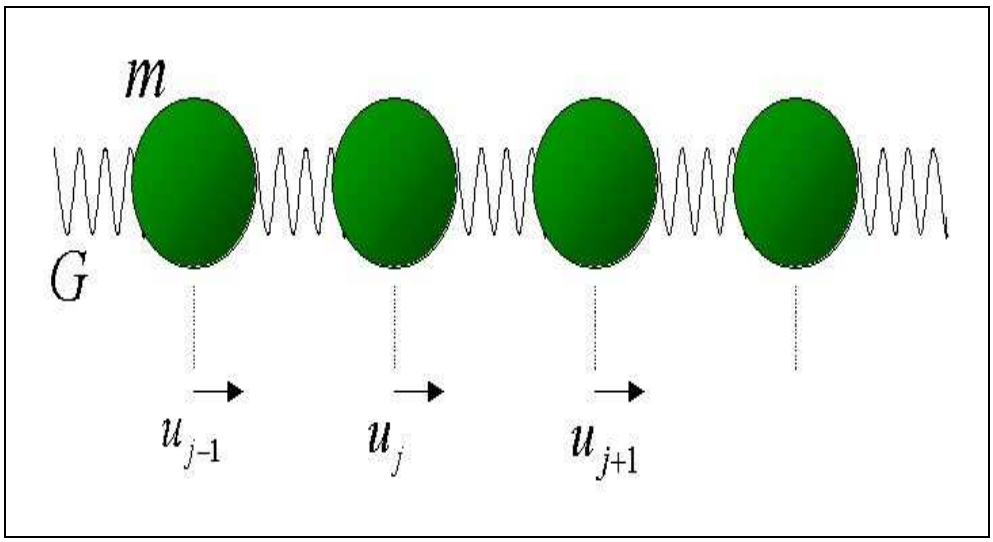

Figure 3.2: A chain of N atoms in one dimension connected by springs.

system as a whole. While we observed the equations of motion of the particles are coupled it is possible to define new vaiables $q_a = u_1 + u_2$ and $q_b = u_1 - u_2$. It can easily be shown that substituting these expressions into the equations of motion produces two simpler equations of the form:

$$\ddot{q}_a + \omega_a^2 q_a = 0, \qquad \ddot{q}_b + \omega_b^2 q_b = 0 \tag{3.8}$$

which are clearly just uncoupled oscillations at the angular frequencies ω_a and ω_b. For an arbitrary disturbance of the system we can represent the resulting motion of the two particles as a superposition of the two normal modes.

This simple example demonstrates the general principle:
the normal modes of a harmonic system do not interact. We will now follow the same analysis for vibrations in solids, limiting ourselves to the case of crystals.

3.2 Crystal vibrations in one-dimension

All the important features of lattice waves in real crystals conveniently emerge from simple one dimensional models. In each case our aim is to find the normal modes.

3.2.1 The monatomic chain

A simple model of a crystal with a single atom basis is shown in Fig. 3.2. We have N balls of mass M connected by springs with spring constant G. The equilibrium separation is a and each ball is labelled by an index j. If the displacements of the atoms from their equilibrium positions are as defined in the figure, then the forces on atom j are $G(u_{j+1} - u_j)$ to the right and $G(u_j - u_{j-1})$ to the left. Writing out Newton's second law, we obtain

$$M\ddot{u}_j = G(u_{j+1} - 2u_j + u_{j-1}). \tag{3.9}$$

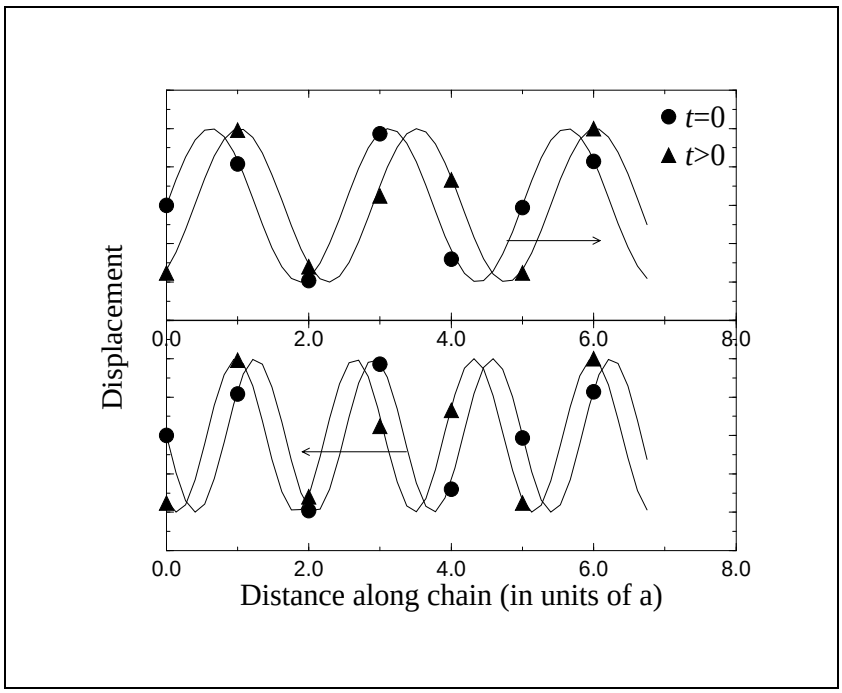

Figure 3.3: Displacements of atoms in a one dimensional chain (plotted as transverse waves for ease of viewing) for $k = 0.8\pi/a$ and $k = -1.2\pi/a$, where a is the length of the unit cell. The arrows indicate the direction of travel of each wave.

The maths is easier if all atoms are equivalent, but we only have a finite number of atoms so we must decide what to do about the atoms on the ends of the chain. The most convenient course of action is to use periodic boundary conditions: $u_{j+N} = u_j$, which amounts to connecting the ends of the chain together. We now have N coupled ordinary differential equations like Eq. 3.9, one for each atom, but they all have the same form.

As before, in a normal mode all the atoms move with the same angular frequency and their vibration amplitudes and phases have a special relationship. You may have already seen elegant matrix methods for decoupling these kinds of equations, but here we are going to anticipate the answer by looking for wavelike solutions with the form:

$$u_j \propto e^{i(kx_j^0 - \omega t)} = e^{i(kja - \omega t)} \tag{3.10}$$

where $x_j^0 = ja$ is the position of the j^{th} atom when none of the springs are compressed or stretched. Notice that this wave is a little strange in that it is only defined at positions corresponding to the undisplaced atomic sites.

Applying the periodic boundary condition to Eq. 3.10 we find that

$$e^{ikNa} = 1 \qquad \Rightarrow \qquad k = \frac{2\pi}{Na} \times \text{integer} = \frac{2\pi}{L} \times \text{integer} \tag{3.11}$$

where L is the length of the entire chain. Clearly if we add $2\pi/a$ (notice that this is a reciprocal lattice "vector" for the one dimension Bravais lattice) onto k

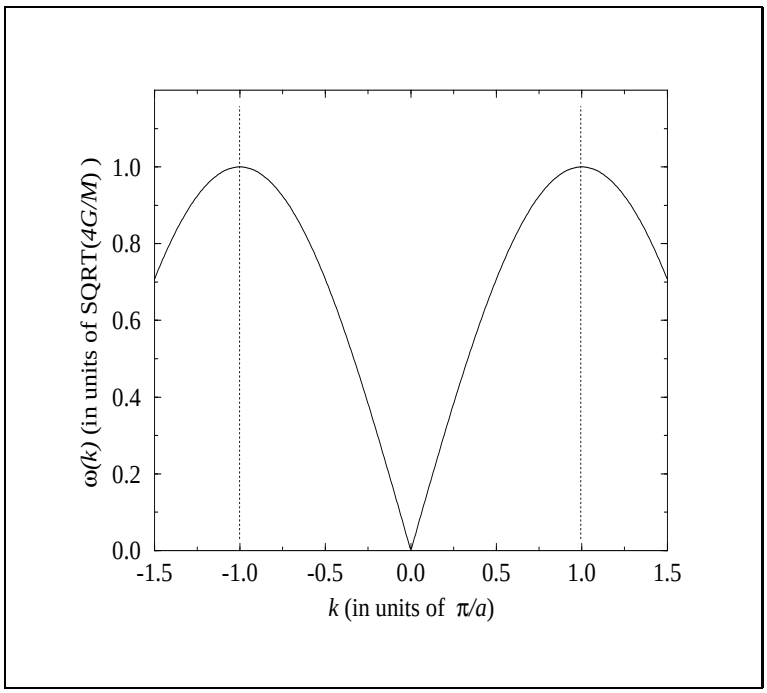

Figure 3.4: Dispersion relation for the normal modes of a harmonic 1D monatomic chain including only nearest neighbour interactions. $\omega(k)$ is periodic in k with period equal to the width of the first Brillouin zone.

then the displacement given by Eq. 3.10 is unaffected. This is illustrated by the displacements plotted in Fig. 3.3 for waves with $k = 0.8\pi/L$ and $k = -1.2\pi/L$. It follows that there are only N distinct allowed values of k and these can be taken between $-\pi/a < k \leqslant \pi/a$, which is the first Brillouin zone.

Substituting Eq. 3.10 into Eq. 3.9 we obtain

$$-\omega^2 M = \left(e^{ika} + e^{-ika} - 2\right) G = \left[\cos(ka) - 1\right] 2G \tag{3.12}$$

$$\Rightarrow \qquad \omega^2 = \frac{4G}{M} \sin^2(ka/2) \tag{3.13}$$

which is the *dispersion relation* for the normal modes, shown in Fig. 3.4. For each value of k there is a normal mode with a certain frequency. A number of features are worthy of comment:

1. In the long wavelength limit ($\lambda \gg a$, i.e. $k \ll 2\pi/a$), ω is linear in k and adjacent unit cells move in phase; the classical limit is correctly reproduced.[2]
2. Dispersion (non-linearity of ω in k) is observed when the wavelength approaches the atomic dimensions. This is characteristic of discrete media.
3. There is a maximum allowed ω, called the *cut-off frequency.*
4. The dispersion relation is periodic in k with period $2\pi/a$, as we predicted.
5. The group velocity $d\omega/dk$ is zero at the Brillouin zone boundaries. Adjacent

[2]Since $\rho = M/a$ and $C = Ga$ for the 1D chain, the wave velocity is (C/ρ) in the long wavelength limit, as it should be.

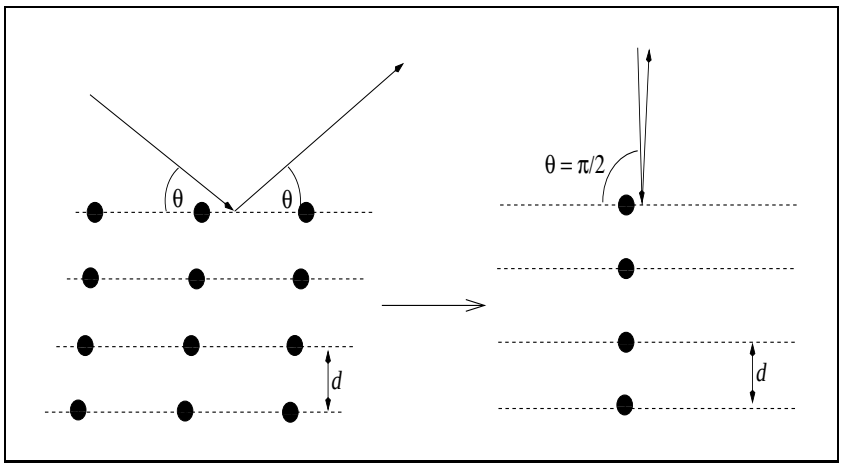

Figure 3.5: The Bragg diffraction condition in one dimension.

unit cells move out of phase. To elucidate this result, consider the Bragg condition for diffraction. If we set $\theta = \pi/2$ and identify the plane spacing shown in Fig. 3.5 with a, diffraction is expected when $k = n\pi/a$. Thus we have a standing wave produced by the interference of lattice waves within the crystal, and so the group velocity must vanish.

So far we have considered only nearest neighbour couplings. Accounting for more distant interactions changes the detailed form of the dispersion relation but all the qualitative features discussed above remain.

It is important to remember that if the monatomic chain is distrurbed from equilibrium then the atoms will start to vibrate in a non-simple way, i.e. in general the disturbance won't correspond to one of the normal modes of the system. However we can express an arbitrary disturbance of the chain w_j, where as before the j subscript labels the atoms, as a superposition of normal modes $\sum_k A_k u_j$ since the chain has N degrees of freedom and there are precisely this many independent normal modes.

3.2.2 The diatomic chain

The diatomic chain is shown in Fig. 3.6. It consists of N unit cells each containing an atom of mass M_1 and one of mass M_2. We assume again that each atom is connected to its nearest neighbours by harmonic springs of strength G, but the interatomic spacing is now $a/2$. If $u_{j,\beta}$ is the displacement of the β^{th} atom in the j^{th} unit cell, then the equations of motion are

$$M_1 \ddot{u}_{j,1} = G(u_{j,2} - 2u_{j,1} + u_{j-1,2})$$

$$M_2 \ddot{u}_{j,2} = G(u_{j+1,1} - 2u_{j,2} + u_{j,1}). \tag{3.14}$$

For each unit cell we have two equations of this form, giving a total of $2N$ coupled ordinary differential equations. As before we anticipate wavelike solutions:

$$u_{j,1} = e^{i(kja-\omega t)}, \qquad u_{j,2} = \alpha e^{i(kja-\omega t)} \tag{3.15}$$

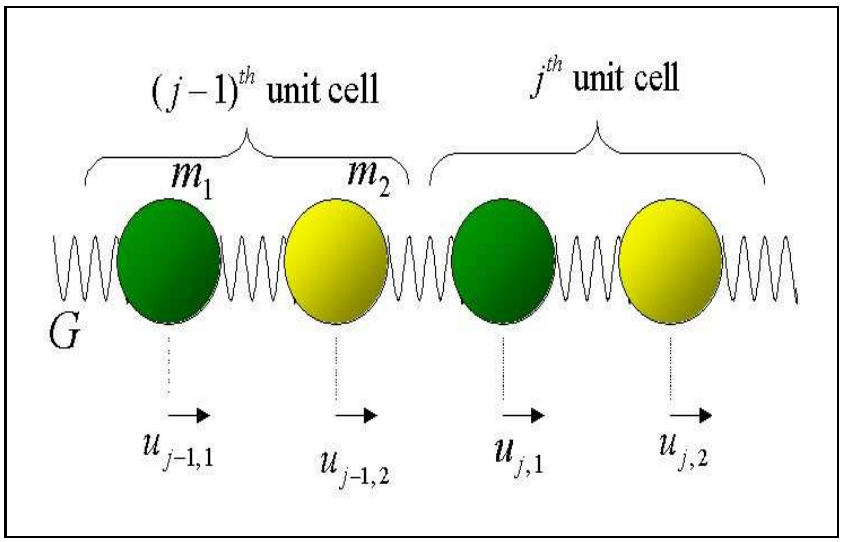

Figure 3.6: The 1D diatomic chain.

where α is a complex quantity allowing a different phase and amplitude for $u_{j,1}$ and $u_{j,2}$.[3] The equations of motion now become

$$\begin{pmatrix} \omega^2 - 2G/M_1 & (1+e^{-ika})G/M_1 \\ (1+e^{ika})G/M_1 & \omega^2 M_2/M_1 - 2G/M_1 \end{pmatrix} \begin{pmatrix} 1 \\ \alpha \end{pmatrix} = 0. \tag{3.16}$$

To obtain a non-trivial solution we require the determinant of the matrix to vanish, and it follows that

$$\omega(k)^2 = \frac{G(M_1+M_2)}{M_1 M_2} \pm G\sqrt{\left(\frac{M_1+M_2}{M_1 M_2}\right)^2 - \frac{4}{M_1 M_2}\sin^2(ka/2)}. \tag{3.17}$$

Again we find periodicity in k with a period of $2\pi/a$, so we need only consider the first Brillouin zone, but on account of the $\pm$ sign we now have two normal mode frequencies for each allowed k, giving a total of $2N$ distinct modes, as required.

The dispersion relation for the diatomic chain is shown in Fig. 3.7. The two *branches* have quite distinct character. The lower frequency branch is non-dispersive at low k (i.e. $\omega \propto k$) and α equals 1, meaning all atoms are moving in phase. This corresponds to the propagation of sound and accordingly the lower branch is designated *acoustic*. For the upper branch we find for $k = 0$ that $\alpha = -M_1/M_2$ and hence the two masses in the unit cell vibrate out of phase with each other and with amplitude which scales inversely with their masses. In ionic crystals this kind of motion produces an oscillating electric dipole. You will probably see in "Atomic and Molecular Physics" that an oscillating electric dipole leads to strong coupling to electromagnetic radiation. For this reason the upper branch is known as the *optical branch*. Resonant vibrations are created in the solid by incident electromagnetic waves with angular frequency equal to $\sqrt{\frac{2G(M_1+M_2)}{M_1 M_2}}$. Since typical angular frequencies for the normal modes in solids

[3] But the frequency ω is common to the whole system for a normal mode.

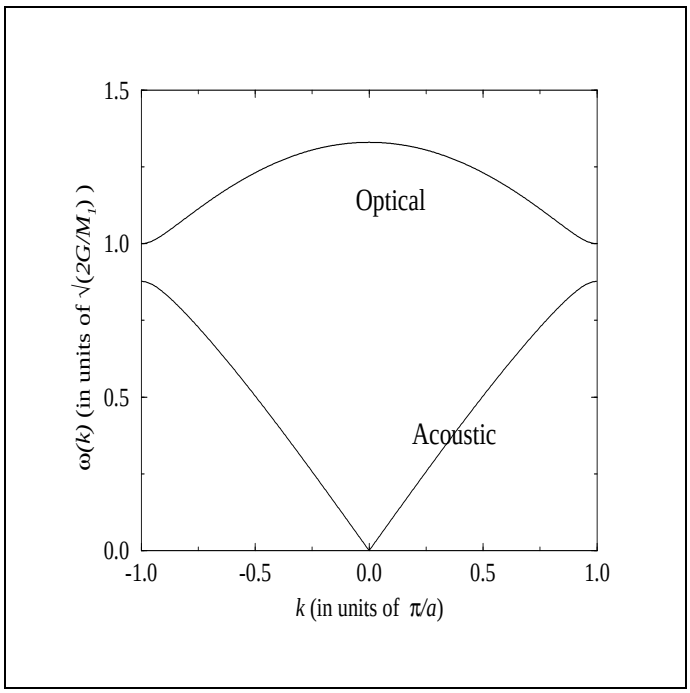

Figure 3.7: Dispersion relation for the 1D diatomic chain including only nearest neighbour interactions, assuming $M_2 = 1.3M_1$.

are of order 10^{13} rads/sec, this leads to a sharp increase in reflectivity of ionic solids in the infra-red region of the electromagnetic spectrum.

3.3 Three dimensions

The generalization to 3D is best achieved using a notation-intensive matrix formulation. There are no real surprises. For solids with one atom per unit cell there can be one longitudinal and two transverse normal modes, possibly with different speeds. If there are p atoms per unit cell there are 3 acoustic branches (one longitudinal and two transverse) and $3(p-1)$ optical branches ($p-1$ longitudinal and $2p-2$ transverse).

The important thing is that we get accustomed to seeing how dispersion relations are represented for 3D crystals. This is a bit tricky at first since the Brillouin zone is now a 3D volume, usually with a strange shape. The way we proceed is by plotting ω vs k along certain directions in the Brillouin zone, as shown in Fig. 3.8. You will have noticed that $\omega(k)$ in 1D is the same for positive and negative k. In 3D many other symmetries become apparent so it is conventional to only bother plotting dispersion relations over an *irreducible wedge* of the Brillouin zone (i.e. we don't bother repeating stuff which is the same).

The normal mode frequencies for a few more crystals are shown in Fig. 3.9. A number of features are apparent:
1. The vibrational frequencies appear to increase along the series argon to lead

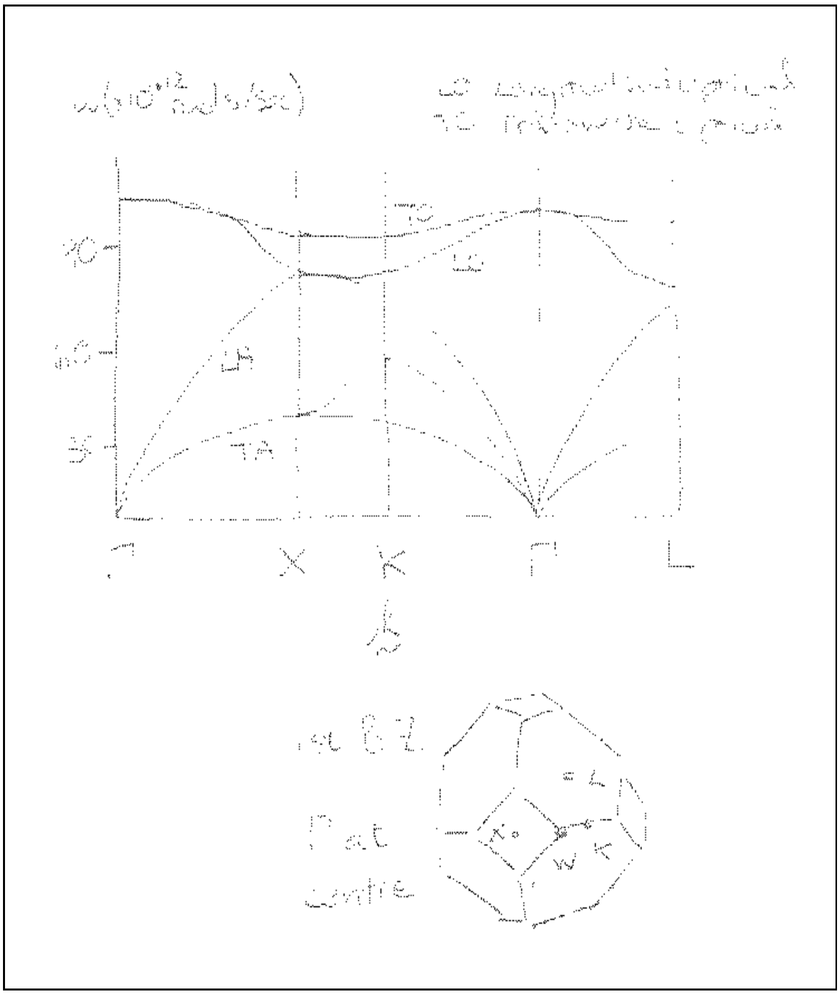

Figure 3.8: The dispersion relation along high symmetry points of the BZ of silicon. The crystal structure of silicon is most simply expressed as a face-centred cubic Bravais lattice with a two atom basis. The reciprocal lattice is therefore a (suitably scaled) body centred cubic lattice and the first BZ is the truncated cubo-octahedron sketched above. Since there are two atoms per primitive unit cell there are acoustic and optical branches.

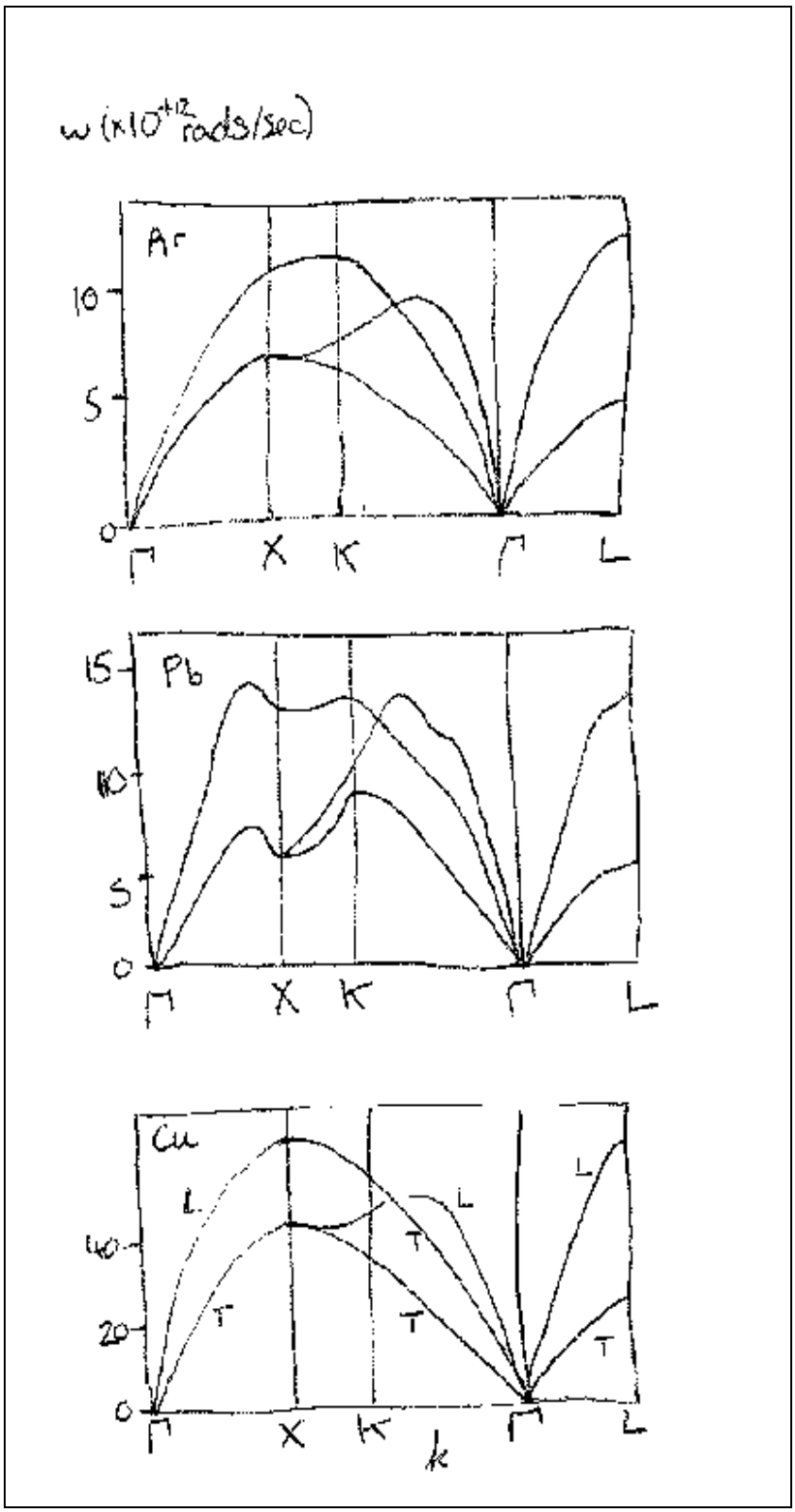

Figure 3.9: The vibrational dispersion relations for Argon, Lead and Copper. Each of these materials crystallises with the face centred cubic crystal structure.

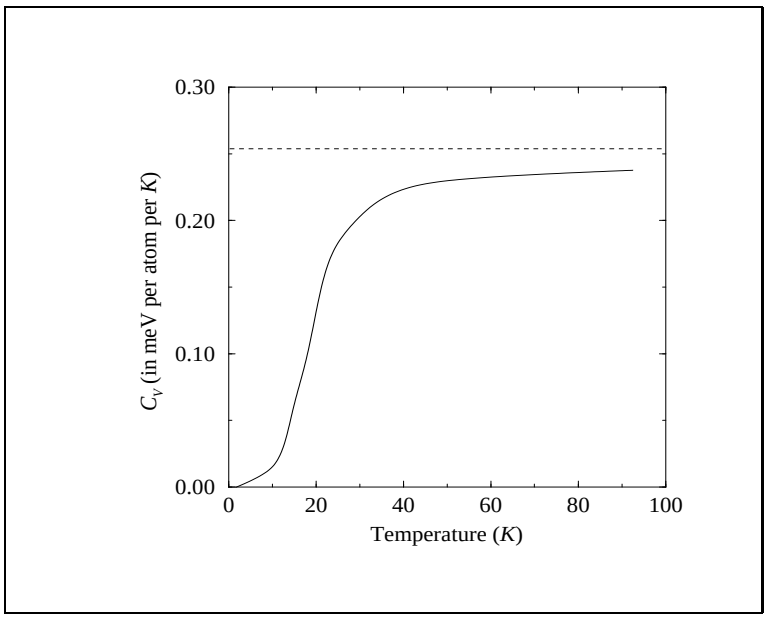

Figure 3.10: Heat capacity of solid Xe. The dashed line indicates the Dulong Petit result.

to copper. From our linear ball and spring equations we can interpret this in terms of an increase in the spring stiffness along the series. While we are not too familiar with the properties of solid argon we can happily accept that copper is more rigid than lead. Looking back to Fig. 3.8 we see that silicon is more rigid still (its vibrational frequencies are even higher) which also makes sense.
2. Along certain directions the transverse modes are doublely degenerate due to the high symmetry of the Bravais lattice.
3. While crystals of silicon, argon and the metals lead and copper are held together by quite different chemical bonds their vibrational spectra are surprisingly similar (except that silicon, having a two atom basis, has optical branches). This justifies our qualitative considerations of balls and springs.

We will soon see how the vibrational modes of real materials can be discovered experimentally.

3.4 Normal modes and Phonons

Earlier we stated that the energy of a crystal above its ground state energy is given by the sum of the energies of its normal modes. Since each normal mode represents an independent oscillator (having both kinetic and potential energy), the equipartition theorem states that the energy density of the crystal at temperature T will be $3Nk_bT$ above its ground state value on account of thermal excitation of the normal modes, where here N is the number of atoms per unit volume. This is just what we need to make sense of the observed heat capacities of most solids measured at room temperature. This is the Dulong-Petit law, known since the middle of the nineteenth century.

Comparison of the Dulong-Petit result and the experimentally measured

heat capacity of Xe is shown in Fig. 3.10. There is disappointing news here. The most obvious difficulty is the apparent collapse of the heat capacity at low temperature. It looks like we are totally missing an important concept here. At high temperature things are much better, but on close inspection the Dulong-Petit limit does not appear to be exactly reproduced by experiment. In the face of these observations we must re-assess the validity of any assumptions or approximations we have made.

The Dulong-Petit result does not require any restriction to only nearest neighbour interactions. So long as only harmonic forces are present, the dynamics of an N particle solid can be exactly mapped onto the dynamics of N independent harmonic oscillators. Here we have our clues - the problems lie with the assumption of harmonic forces and with the type of dynamics we used (i.e. Newtonian).

In our analysis of the normal modes of vibration in crystals we derived the relation between angular frequency ω and wavevector $\underline{k}$, but at no point did we explicitly consider the *amplitude* of the vibrations. Classical dynamics says that the energy of a harmonically oscillating mass is $Mr^2\omega^2$ where r is the vibration amplitude. Thus the average square amplitude of thermally excited vibrations should be proprtional to k_bT. At low T we have small amplitude vibrations and so Taylor's theorem tells us the harmonic approximation should work rather well.[4] So the harmonic approximation cannot explain the low T problem, but it probably has something to do with the high T problem since true interatomic potentials are not really harmonic and this becomes apparent at high T.[5]

The low T heat capacity problem signals a rather fundamental problem - in fact the complete breakdown of classical physics. Just as in the "black-body problem" of electromagnetism, the normal mode oscillators of a harmonic crystal can only vibrate with certain allowed amplitudes such that their energy complies with the Planck hypothesis:

$$E = (j + 1/2)\hbar\omega \qquad j = 0, 1, 2, 3, \cdots . \tag{3.18}$$

This simple expression brings a number of thoughts to mind:

(i) A crystal cannot have zero vibrational energy. There is at least $\hbar\omega/2$ of *zero-point energy* in each mode. (We have already appealed to this phenomenon when we discussed the cohesive energy of the noble gas solids.)
(ii) Each normal mode can only gain or lose energy in units of $\hbar\omega$.
(iii) We can adopt similar quantum terminology as used in electromagnetism. Instead of referring to an oscillator in branch s with wavevector $\underline{k}$ being in its j^{th} excited state, we say there are j *phonons* of type s with wavevector $\underline{k}$ in the crystal. When an oscillator is converted from its j_1^{th} to its j_2^{th} excited state, we say that $j_1 - j_2$ phonons have been emitted or absorbed by the oscillator.
(iv) One can see that the thermal properties of solids will depend on the relation of k_bT to $\hbar\omega$. If k_bT is greater than $\hbar$ times the typical normal mode frequency of a crystal then their quantised nature should not be apparent. On the other hand, if it is smaller, then a quantum theory will be essential.

[4] In fact the harmonic approximation is exact in the $T \to 0$ limit.

[5] Generally it is harder to push atoms together than it is to pull them apart. Whatever their precise nature, interatomic (or interionic) attractive forces are always longer range than the Pauli repulsive force so the potential energy curve is not a parabolic function of interatomic separation.

From Fig. 3.10 one can conclude that the transition from the quantum to the classical regime occurs at ~ 50 K.[6] This implies that the characteristic angular frequency of lattice vibrations (in xenon at least) is of order 10^{12} rads/sec, corresponding to a phonon energy of just a few milli-electron volts. If we were to have plotted C_V against T for a different solid, say copper or diamond, then we would have seen almost exactly the same shape of curve. This is another example of the fact that vibrations in crystals are much of a muchness. It seem s like we can characterise the vibrational properties of a particular solid by simply one number: the characteristic vibrational frequency, or equivalently a characteristic temperature.

3.5 The heat capacity of crystals

3.5.1 Thermal energy of a single quantum harmonic oscillator

Consider a single quantum harmonic oscillator in thermal equilibrium with a heat reservoir at temperature T. We know that the energies of the oscillator in its j^{th} excited state are $(j+1/2)\hbar\omega$. We don't know which state the oscillator will be in, but the Boltzmann law tells us that the *probability* P_j that the oscillator will be in its j^{th} excited state is

$$P_j \propto e^{-E_j/(k_b T)}. \tag{3.19}$$

The constant of proportionality can be determined by the fact that the sum of all these probabilities must be 1, leading to

$$P_j = e^{-j\hbar\omega/(k_b T)} \left(1 - e^{-\hbar\omega/(k_b T)}\right). \tag{3.20}$$

The average energy of the single oscillator is therefore

$$\langle E \rangle = \sum_{j=0}^{\infty} P_j E_j = \hbar\omega \left(\frac{1}{2} + \frac{1}{e^{\hbar\omega/(k_b T)} - 1}\right). \tag{3.21}$$

This expression is very similar in form to the equation for the excited states of the oscillator. We can therefore deduce that the average value of the quantum number j for an oscillator in thermal equilibrium with a heat bath is

$$\langle j \rangle = \frac{1}{e^{\hbar\omega/(k_b T)} - 1}. \tag{3.22}$$

We can interpret the energy $\langle E \rangle$ as comprising a zero point contribution $\hbar\omega/2$ and a contribution $\hbar\omega$ for each of the $\langle j \rangle$ phonons that are present (on average). You will recognise this expression as the Bose-Einstein distribution function f^{ω}_{BE} which gives the number of *bosons* with energy $\hbar\omega$ in equilibrium at temperature T.

[6] The equivalent temperature in the free electron gas theory was the "Fermi tempertaure" $T_f = \epsilon_f/k_b$. We saw that this is many thousands of degrees Kelvin and so electrons are firmly in the quantum regime for most purposes. This is not so for phonons.

3.5.2 The Einstein model

When Planck made the crucial breakthrough in explaining the black-body "UV-catastrophe", Einstein realised that the same basic idea of quantised oscillators was also relevant to the heat capacity problem of solids.[7] Einstein assumed that the vibrational properties of a crystal of N atoms could be modelled by $3N$ independent harmonic oscillators with angular frequency ω_E, known as the *Einstein frequency*. The energy density of the solid is then given by

$$U = \frac{3N\hbar\omega_E}{W}\left(\frac{1}{2} + \frac{1}{e^{\hbar\omega_E/(k_bT)} - 1}\right) \tag{3.23}$$

where W is the volume of the crystal, and so the heat capacity[8] is

$$C_V = \left(\frac{\partial U}{\partial T}\right) = \frac{3Nk_b}{W}\left(\frac{\hbar\omega_E}{k_bT}\right)^2 \frac{e^{\hbar\omega_E/(k_bT)}}{\left(e^{\hbar\omega_E/(k_bT)} - 1\right)^2}. \tag{3.24}$$

Consider first the high T regime, by which we mean $k_bT \gg \hbar\omega_E$. To simplify the algebra it is conventional to define the Einstein temperature Θ_E:

$$\Theta_E = \frac{\hbar\omega_E}{k_b} \tag{3.25}$$

and so "high temperature" means $T \gg \Theta_E$. We can write the Bose-Einstein function as

$$\langle j\rangle = \frac{T}{\Theta_E}\left(1 - \frac{1}{2}\frac{\Theta_E}{T} + \cdots\right) \tag{3.26}$$

and so the energy density of the crystal becomes

$$U \to \frac{3Nk_b\Theta_E}{W}\left(\frac{1}{2} + \frac{T}{\Theta_E} - \frac{1}{2}\right) = \frac{3Nk_bT}{W}. \tag{3.27}$$

Thus the high temperature limit of the Einstein model reproduces the Dulong-Petit heat capacity, i.e. $C_V \to 3Nk_b/W$.

The variation of the Einstein heat capacity with T is plotted in Fig. 3.11. Comparison with experimental data for Xe shown earlier strongly suggests that the Einstein model has corrected the most flagrant failures of the classical theory. The Dulong-Petit limit is reproduced and C_V collapses to zero at low temperature. The physics of the low T region can be put into words: when $T \ll \Theta_E$ the available thermal energy is not sufficient to excite (with appreciable probability) the oscillators from their ground state. A small increase in T in this region increases the likelihood of excitation only a negligible amount and so the energy density of the solid increases negligibly, i.e. the heat capacity vanishes in the low T limit.

Closer inspection of Eq. 3.24 shows that in the Einstein model C_V vanishes exponentially as T tends to zero. On the contrary we saw in Chapter 1 that at low temperature the (vibrational contribution to the) heat capacity varies as T^3. The Einstein model contains the essential physics, but it gets the detail wrong. The problem is not difficult to spot - we have assumed that the crystal consists of independent harmonic oscillators all with the same frequency, ω_E. We know that this is not true.

[7] Although we haven't got time to go into it, we should also mention that similar problems existed in explaining the heat capacities of molecular liquids.

[8] In the harmonic approximation $C_P = C_V$, but don't worry about that.

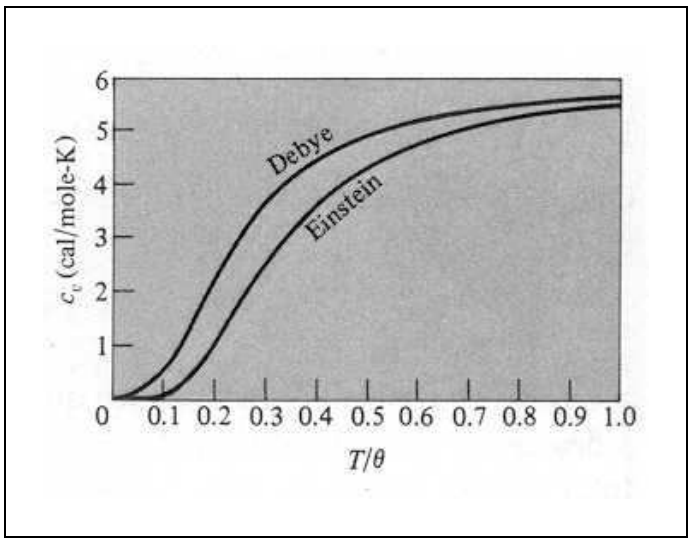

Figure 3.11: Heat capacity of a crystal in the Einstein and Debye models.

3.5.3 The vibrational density of states (modes)

In sections 3.1-3.3 we saw that crystals can vibrate with a spectrum of frequencies. We must now graft together our classical discussion of the normal mode frequencies of a crystal and the quantum theory of the heat capacity. The simplest way to proceed is by introduction of the *density of vibrational states.*

The normal mode solutions of the equations of motion of the vibrating harmonic crystal are wave-like, each mode having a particular wavevector $\underline{k}$. The mathematics of counting these modes is very similar to that for counting the states in the wave-like solutions of the Schrödinger equation for the free electron gas from Chapter 1. In particular, we will need to figure out (i) how many states (i.e. modes) are allowed, and (ii) what is the probability that these states are occupied (i.e. how many phonons are there).

The application of periodic boundary conditions for a crystal with volume W leads to a density of allowed phonon wavevectors given by

$$g^{\underline{k}} = \frac{W}{(2\pi)^3}. \tag{3.28}$$

There is an important difference however - we need only consider $\underline{k}$ to lie within the first Brillouin zone. The definition of $g^{\underline{k}}$ requires $g^{\underline{k}}\delta\underline{k}$ to be the number of allowed modes with wavevector in the range $\underline{k}$ to $\underline{k} + \delta\underline{k}$. Similarly, $g^{\omega}\delta\omega$ gives the number of allowed modes with angular frequency in the range ω to $\omega + \delta\omega$. Given a complete knowledge of the dispersion curves $\omega_s(\underline{k})$ of a crystal then we can deduce g^{ω}. This is shown in Fig. 3.12 for copper. Notice that peaks in g^{ω} are observed whenever the branches are flat since this ensures there are many modes in a small energy range.

If we are to calculate the energy density of a crystal at temperature T we must now supplement our knowledge of the allowed modes of a crystal by specifying the extent that these modes will be excited. In other words, we must specify how many phonons of type s (i.e. in branch s), with wavevector $\underline{k}$ and

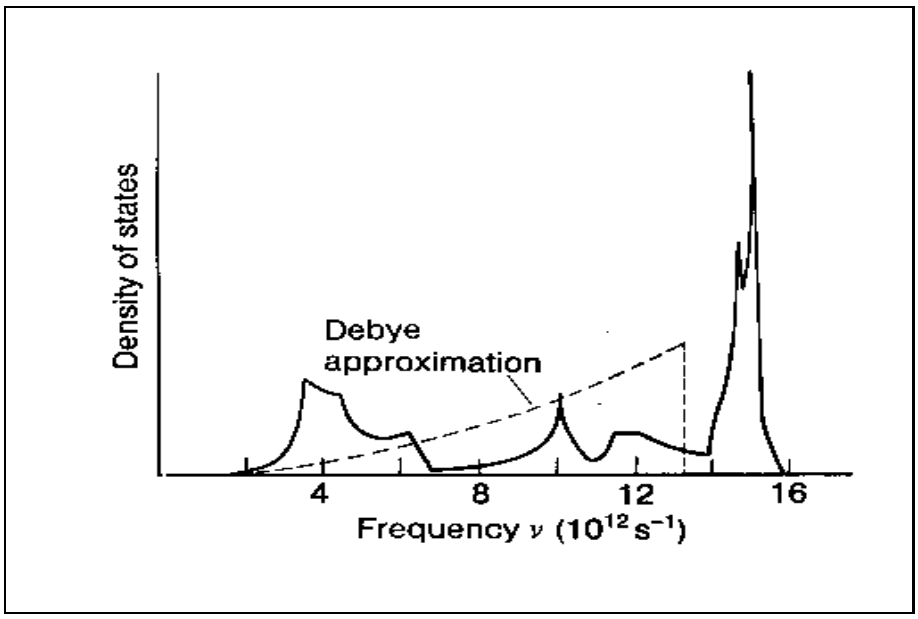

Figure 3.12: The vibrational density of allowed states g^ω for silicon. The dashed curve is the Debye approximation.

energy $\hbar\omega$ will be present at temperature T. This is given by the Bose-Einstein function. Finally, the vibrational energy density of the solid can now be written

$$U = \int g^\omega \, f_{BE}^\omega \, \hbar\omega \, d\omega = \int g^\omega \left[\frac{1}{2} + \frac{1}{e^{\hbar\omega/(k_b T)} - 1}\right] \hbar\omega \, d\omega. \tag{3.29}$$

3.5.4 The Debye model

In the previous subsection we showed the formal link between the spectrum of vibrational modes of a crystal and its heat capacity. Plots of $\omega_s(\underline{k})$ throughout the Brillouin zone of a crystal look rather messy and it is hard to interpret what they actually mean. In order to understand the qualitative behaviour of the heat capacity it is enough to consider a simple model for the density of vibrational states. We saw above that the Einstein model, for which $g^\omega \propto \delta(\omega - \omega_E)$, is a bit too simplistic. However, inspection of Fig. 3.2 suggests the approximation $\omega \propto k$, originally proposed by Debye, might be useful. This assumption neglects the dispersion due to the discrete structure of the crystal.

To compute the density of allowed vibrational states for the Debye model we again proceed in the manner seen in Chapter 1 for the free electron gas. First we consider the volume of k-space lying between the spheres with radii k and $k + \delta k$. The number of allowed modes with wavevector within this region is by definition $g^k \, \delta k$. But this is also equal to $g^{\underline{k}} \delta\underline{k}$, where $\delta\underline{k}$ is the chosen k-space volume, and it follows that

$$g^k = \frac{W}{(2\pi)^3} 4\pi k^2. \tag{3.30}$$

To convert g^k to g^ω we play the same trick:

$$g^\omega \, d\omega = g^k \, dk. \tag{3.31}$$

For the Debye model we have $\omega = ck$, where c is something like the speed of sound, and so we obtain

$$g^\omega = \frac{g^k}{d\omega/dk} = \frac{W}{(2\pi)^3}\frac{4\pi k^2}{c} = \frac{W}{2\pi^2}\frac{\omega^2}{c^3}. \tag{3.32}$$

There are a few things we have to do to this expression:
(i) It is conventional to divide by the crystal volume W, so we are really dealing here with a density of states per unit crystal volume.
(ii) We must multiply by a factor of three, since there are three acoustic branches for a three dimensional solid.[9]
(iii) We might worry about what to do about the optical branches for crystals with a multi-atom basis. Fig. 3.5 shows that these are quite flat and so they are often modelled by a single oscillator frequency as in the Einstein model. We could do this, but for simplicity we will side-step the problem by considering only crystals with a one atom basis.

Even with these modifications and qualifications, the density of allowed vibrational states given by Eq. 3.32 is deficient in one important respect. We noted earlier that for a crystal with N unit cells there are N distinct allowed wavevectors. In 3D there are therefore $3N$ allowed modes (for a monatomic crystal). We need to specify a cut-off frequency ω_D (called the *Debye frequency*) such that the integral of Eq. 3.32 yields the correct total number of allowed modes:

$$\frac{3N}{W} = \int_0^{\omega_D} g^\omega \, d\omega = \frac{3}{2\pi^2 c^3}\int_0^{\omega_D} \omega^2 \, d\omega \qquad \Rightarrow \qquad \frac{\omega_D^3}{c^3} = k_D^3 = 6\pi^2\frac{N}{W}. \tag{3.33}$$

The Debye approximation for g^ω of copper is shown in Fig. 3.12.

Clearly ω_D is the characteristic vibrational frequency in the Debye model. It is also useful to define the *Debye temperature* $\Theta_D = \hbar\omega_D/k_b$ and the *Debye wavevector* $k_D = \omega_D/c$. We could view the Debye model as the replacement of the true phonon spectrum of a crystal $\omega_s(\underline{k})$, where $\underline{k}$ has N allowed values in the first Brillouin zone, with the dispersion relation $\omega_s(\underline{k}) = ck$, where $\underline{k}$ has N allowed values within the "Debye sphere" $k < k_D$.

Putting everything together, the Debye expression for the energy density of a monatomic crystal is

$$U = \frac{3\hbar}{2\pi^2 c^3}\int_0^{\omega_D} \omega^3 \left(\frac{1}{2} + \frac{1}{e^{\hbar\omega/(k_b T)} - 1}\right) d\omega. \tag{3.34}$$

Differentiating this expression to obtain the heat capacity yields the result shown in Fig. 3.11. Again the Dulong-Petit limit is attained for high T (i.e. $T \gg \Theta_D$), but what about the low T limit? Making the substitution $x = \hbar\omega/(k_b T)$ we obtain

$$C_V = \frac{9Nk_b}{W}\left(\frac{T}{\Theta_D}\right)^3 \int_0^{\Theta_D/T} \frac{x^4 e^x}{(e^x - 1)^2} dx. \tag{3.35}$$

[9] Strictly speaking we should be taking the constant c as an average for the three acoustic branches, but this is an annoying detail.

At low T we can take the upper limit of integration to be infinite, whereupon the integral takes the value $4\pi^4/15$. At last we are left with the desired T-cubed behaviour at low temperature.

We saw above that within the Debye model, the energy density of a crystal at a given T is determined by a single parameter, namely the Debye temperature of the material Θ_D. One is entitled to ask whether this is reasonable. Fig. 3.12 shows that while the Debye approximation misses some of the details it gets the essential features of the phonon spectrum i.e. the correct behaviour near $\omega \sim 0$, a reasonable upper bound and (by construction) the correct total number of modes. Since the vibrational density of states only enters the expression for the energy density in integral form (see Eq. 3.30) the precise details get washed out.

The Debye temperatures for aluminium, lead, neon and diamond are 394, 88, 63 and 1860 K. What do these numbers mean? The frequency of a harmonic oscillator is $\sqrt{G/M}$ and so we would expect that heavy atoms would have lower frequencies than light ones. The comparison of aluminium and lead bears out this expectation. Both these elements are metallic (and so one might expect a similar "spring constant"), but lead atoms are much heavier than aluminium atoms. Equating the G with the second derivative of the energy with respect to atomic displacement, one expects that crystals with high cohesive energies will have a high "spring constant" and hence a high Debye temperature. Diamond is a prime example of this, while neon demonstrates the converse. In summary, solids with a high Θ_D tend to have a high cohesive energy, to be rigid, and incompressible.

The Debye temperature in lattice dynamics plays a similar role to the Fermi temperature in the free electron gas theory. Both separate regions where quantum statistics apply from regions where classical statistics are valid. Since the Fermi temperature is so high ($\sim 10^5$ K) we only encounter the quantum mechanical regime in free electron theory. For most solids T_D is of order 10^2 K and so both classical and quantum regimes of lattice dynamics are readily accessible.

3.6 Diffraction from a vibrating crystal

In this section we are considering again *scattering*, i.e. the probe is not absorbed[10] but rather survives interaction with matter, possibly suffering some modification in the process.

Consider the following expression:

$$A'(\underline{q}) \propto e^{-i\omega_0 t} \sum_j e^{-i\underline{q}.(\underline{R}_j + \underline{u}_j(t))}. \tag{3.36}$$

This is Eq. 2.34 (which gives the amplitude of waves scattered by a solid) for a crystal with a single atom basis but with three modifications. We have explicitly included the factor $\exp(-i\omega_0 t)$ arising from the time dependence of the incident wave, suppressed the atomic form factor, and allowed the atomic coordinates

[10] When we looked the optical branches of crystals with a two atom basis we saw that it is possible for a photon to be converted to a phonon and *vice versa* provided there is an oscillating electric dipole in the corresponding normal mode vibration. Energy and "crystal momentum" must also be conserved, as we will see below. It is also possible for a photon to annihilate in the production of multiple phonons.

to vary with time. If we now approximate $\exp(-i\underline{q}.\underline{u}(t))$ using the binomial expansion to first order, and if we write the displacements $\underline{u}_j(t)$ in the form used in Eq. 3.10, we find that the intensity of the Bragg peaks is reduced by a factor $\exp(-q^2\langle u^2\rangle/3)$, where q is the magnitude of scattering vector and the term in angled brackets denotes the mean square displacement. The intensity removed from the Bragg spots by this "Debye-Waller" factor turns up as a diffuse background.

After the leading (elastic) term we obtain an inelastic term of the form

$$e^{-i(\omega_0 \pm \omega(\underline{k}))t} \sum_j e^{i(\underline{q} \pm \underline{k}).\underline{R}_j}. \tag{3.37}$$

While our scattering theory introduced in Chapter 2 was semi-classical, there is an obvious quantum mechanical interpretation of the term above: when scattered from a crystal incident particles can absorb/create a phonon with energy $\hbar\omega(\underline{k})$ and wavevector $\underline{k}$.

For liquids, where $\underline{u}(t)$ is not wavelike, the spectrum of inelastically scattered waves can also yield information on molecular dynamics. We haven't got time to got into this however.

3.6.1 Conservation laws and "crystal momentum"

The elastic scattering considered in the previous chapter, for which the energy of the probe (which could be photons or neutrons) is conserved:

$$E'_{\text{probe}} = E_{\text{probe}} \tag{3.38}$$

can be considered as *zero-phonon scattering*. (Here we add the "probe" subscript to distinguish between the energy and wavevector of the neutrons or photons and those of the phonons.) In addition we have the von Laue diffraction condition:

$$\underline{k'}_{\text{probe}} = \underline{k}_{\text{probe}} + \underline{K} \tag{3.39}$$

although we did not think of this as a conservation rule. When a phonon is destroyed (positive sign) or created (negative sign) when a particle[11] is scattered in a crystal these equations become

$$E'_{\text{probe}} = E_{\text{probe}} \pm \hbar\omega_s(\underline{k}) \tag{3.40}$$

$$\hbar\underline{k}'_{\text{probe}} = \hbar\underline{k}_{\text{probe}} \pm \hbar\underline{k} + \hbar\underline{K}. \tag{3.41}$$

If we regard $\hbar\underline{k}$ as the momentum carried by the phonon, this last expression looks very much like the conservation of linear momentum.

This is not quite correct however. Consider a phonon travelling through the crystal before the diffraction experiment begins. Momentum is only a constant of motion when a particle moves through a constant potential. If a ball rolls downhill or across the corrugations of a sheet of corrugated iron it is obvious that the momentum of the ball is not a constant of its motion. The potential

[11] The single scattering theory we developed in Chapter 2 is suited to the description of neutron or x-ray scattering.

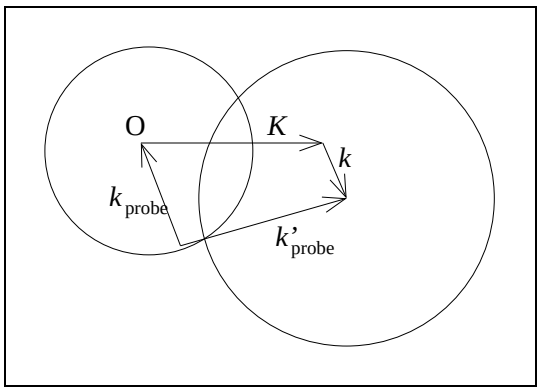

Figure 3.13: Schematic representation of the conservation rules in the diffraction of neutrons by a crystal. The right hand circle is centred on $\underline{K} + \underline{k}$ and has radius determined by energy conservation.

within a crystal is not constant, and so its stationary states are not momentum eigenfunctions and in general $\hbar\underline{k}$ does not correspond to true kinematic momentum.[12]

But the potential in a crystal is *periodic*, and for every symmetry there is a corresponding conservation law. In periodic systems $\hbar\underline{k}$ is conserved, but with an uncertainty of $\hbar\underline{K}$, where $\underline{K}$ is a reciprocal lattice vector. Since $\hbar\underline{k}$ plays a similar role in periodic systems as momentum plays when the potential is constant, we often refer to it as the *crystal momentum* of the phonon. The conservation of crystal momentum makes sense because we have already seen that a phonon in branch s and with wavevector $\underline{k} + \underline{K}$ is indistinguishable from a phonon in branch s and with wavevector $\underline{k}$.

3.6.2 Phonon spectroscopy

It should now be clear that measuring the *inelastic* scattering from a crystal will allow its phonon spectrum to be investigated. X-ray scattering is not well suited for this purpose however since phonon energies are of order 10^{-3} eV while the photon energy of x-rays used in diffraction experiments is of order 10^4 eV. It is extremely difficult to distinguish elastic from inelastic scattering in the x-ray regime since it is extremely difficult to create an x-ray beam which is sufficiently monochromatic (i.e. with a sufficiently narrow spread of photon energies) and sufficiently intense. The visible region of the electromagnetic spectrum ($E_{\text{probe}} \sim 1$ eV) is much easier to deal with since lasers, which are both intense and extremely monochromatic, are cheap and readily available. The catch is that the wavevector of these photons is $\sim 10^7$ m^{-1} $= 10^{-3}$ Å^{-1}, which is extremely small on the scale of the Brillouin zone of solids. Thus inelastic scattering in the visible region probes only those phonons with $\underline{k} = 0$ (known as *Brillouin scattering* if an acoustic branch is probed, and *Raman scattering* for the optical branch).

In phonon spectroscopy it is neutrons that steal the show. Neutrons with de Broglie wavelength comparable to the interatomic spacing of solids have en-

[12] We will see a spectacular demonstration of this principal in the next chapter.

ergy of only ~ 0.02 eV, and it is therefore relatively easy to resolve inelastically scattered neutrons. One can proceed as follows:

(i) Choose the wavevector and energy of the incident neutrons, $\underline{k}_{\rm probe}$ and $E_{\rm probe}$. (This means specifying the speed and direction of the neutrons, but we need not won't worry about how this is done, we'll just take it for granted that it can be.)
(ii) Select $\underline{k}$, the wavevector of the phonons we are interested in.
(iii) Select $\underline{q}$, the scattering vector of the neutrons (remember this is just $\underline{k}'_{\rm probe} - \underline{k}_{\rm probe}$). $\underline{q}$ is equal to $\underline{K} + \underline{k}$, where $\underline{K}$ is a particular reciprocal lattice vector of the crystal.
(iv) Measure the energy spectrum of neutrons scattered in the direction defined by $\underline{q}$. For the chosen $\underline{q}$ and arbitrary $\underline{k}_{\rm probe}$, Eq. 3.40 will probably not be satisfied and so no scattered neutrons will be observed for this geometry.
(v) Keeping $E_{\rm probe}$ constant, change $\underline{k}_{\rm probe}$ and measure the neutron spectrum again. When the energy conservation rule is satisfied there will be a peak in the scattered neutron spectrum at energy $E_{\rm probe} \pm \hbar\omega_s(\underline{k})$.

Thus the energies of the observed peaks[13] together with Eq. 3.40 give the phonon energies for the chosen $\underline{k}$. In this way the phonon branches can be mapped throughout the Brillouin zone. The geometry of this method is shown schematically in Fig. 3.13.

Of course neutron diffraction can also be used to look for the elastic Bragg peaks for the purpose of crystal structure determination, just as for x-ray diffraction. Here too neutrons have some advantages. They are scattered primarily by interaction with the nuclei in solids[14] which can be considered point like on the scale of interatomic distances. As a result the atomic form factor for neutron diffraction (i.e. the Fourier transform of the distribution of nucleons in an atom) is roughly independent of scattering vector and so diffraction intensities do not fade away at high q. We have already noted that the scattering strength for neutrons does not vary that much across the periodic table and so neutron diffraction does not have any trouble finding very light atoms, unlike x-ray diffraction.

Given the advantages of neutron diffraction in structure analysis as well as the capability to map out the entire phonon spectrum of a solid, one might wonder why anyone would bother with x-ray diffraction at all. The answer is mainly a matter of practicality. It is rather easy to generate, monochromate (i.e. energy select) and detect x-rays. Neutrons are much harder to come by. While there are dozens of x-ray sources in the University of Edinburgh alone, neutron diffraction can only be performed at extremely expensive custom built national facilities.

3.7 Dynamics in non-crystalline systems

Let's pause briefly to consider the possibility of vibrations in non-crystalline matter. (See also Dr. Egelhaaf's "Macromolecular Physics" lectures next term.)

[13] Experimental results reveal the inelastic peaks, though distinct, to be broadened somehow.
[14] Since neutrons have spin there is also a weak coupling to the magnetic moments of the electrons, but this is only significant in magnetic materials.

Firstly we could wonder what the effect of impurities might have. These might come in the form of chemical impurities, isotopic impurities, or missing atoms for example. When a vibration spreading through a crystal reaches an impurity it will get messed up. The impurity will attempt to vibrate in slightly the wrong way. This is a bit like trying to push a swing at the wrong moment - the amplitude of the swing will probably be degraded. Thus impurities lead to a damping of lattice vibrations, and we might expect a similar effect from other types of defect (by which we mean departure from perfect crystallinity, grain boundaries, surfaces, etc.). There is a very important concept here: *disorder leads to localisation.*

The localisation of a vibration means, according to the uncertainty principle, the wavevector of the wave will be spread out. The more confinement in real space, the more uncertain will be the wavevector. It is then quite natural to associate a phonon of wavevector $\underline{k}$ with a *wavepacket*, i.e. a superposition of waves with wavevectors centred on $\underline{k}$.

In amorphous solids the disorder is more thorough. For longwavelength vibrations (i.e. where the wavelength greatly exceeds the atomic scale) the details of atomic scale geometry are irrelevant and the material behaves as an elastic continuum, supporting well defined modes. For shorter wavelengths (larger k), disorder leads to such a smearing in wavevector that $\Delta k \sim k$ and hence dispersion relations have no validity.

3.8 Anharmonicity

We have seen that the harmonic approximation leads to a surprisingly simple theory of the vibrating lattice. Taylor's theorem reassures us that we are on quite solid ground with this approximation, for small amplitude vibrations at least. Eq. 3.22 on page 77 tells us that as $T \to 0$, $\langle j \rangle \to 0$ and so there are no phonons about.[15] Equivalently, we could say that the normal mode oscillators in a crystal are in their ground states and only zero point motion is present at absolute zero. There must always be a range of temperature which the vibrational amplitudes are sufficiently small for the harmonic approximation to be valid, but outside this region the atoms vibrate sufficiently for them to notice anharmonicity. We have already seen that the failure of C_V to approach the classical (Dulong-Petit) limit is an example of this.

One might think that anharmonicity is an annoying departure from mathematical simplicity which merely messes up the fine details. In fact this is not the case - there are a few quite obvious physical effects which are wholly determined by anharmonicity, thermal expansion coming immediately to mind. A quick discussion of *thermal conduction* illuminates the main points.

The normal modes of a purely harmonic lattice are independent and, once created, a vibration will persist indefinitely. In other words phonons don't interact with each other. Now one can send pulses (i.e. phonons) across a solid and hence transport thermal energy and this is the basis of an explanation of thermal conductivity, but it's incomplete since non-interacting phonons give rise to indefinite thermal currents and hence *infinite* thermal conductivity.[16] However

[15] Note that this behaviour is in stark contrast to what we saw for the electron gas for which the total number of particles is constant.

[16] We observed earlier that impurities and defects could lead to damping of harmonic vibra-

the phonons of a real crystal are not quite true stationary states of the exact Hamiltonian which contains an anharmonic potenial. Experimental observation of single phonon loss/gain peaks in the inelastic neutron scattering[17] suggests that the phonon picture is substantially correct and therefore a good starting point for a perturbation theory treatment. In first order perturbation theory it can be shown that the cubic (and quartic etc.) terms in the Taylor expansion of the interatomic potential give rise to processes in which the occupation number of 3 (4 etc.) modes change. In other words the most important types of phonon-phonon interactions lead to either the decay of one phonon to create two new ones, or the combination of two phonons to form one new one.

A schematic diagram of the former is shown in Fig. 3.14a. Since the principles of energy and crystal momentum do not rely upon the harmonic approximation they must apply here. It follows that the direction and magnitude of phonon momentum is unchanged by the interaction in Fig. 3.14a and hence thermal currents are not degraded by such processes. The collision shown in Fig. 3.14b is significant because the resulting phonon has wavevector lying outside the first BZ. We know that this must correspond precisely to a normal mode with wavevector *inside* the first BZ, shown by the dotted line in the Figure. This is perfectly consistent with the law of crystal momentum which requires $\underline{k}$ to be conserved give or take a reciprocal lattice vector $\underline{K}$. The result of these *Umklapp* collisions (the other type are simply called "normal"), is that the phonon momentum can be reversed and hence thermal currents are degraded. We have discovered the origin of thermal *resistance*.[18]

3.9 Summary

This chapter has explored the properties and consequences of "deformation waves" in solids. In the long wavelength limit (or short wavevector limit) these are just sound waves. We know solids transmit sound because they are "springy", i.e. both stretchable and compressible. Compressibility, rigidity, Hooke's law etc. all belong to the field of continuum elasticity theory, but we found that looking into the atomic origins of these effects leads to some new physics. We saw this by modelling a vibrating crystal by a one dimensional chain of balls connected by harmonic springs. Displacement of an atom causes its neighbours to be displaced and so on down the chain like a wave. The crucial point here is that the deformation of the "solid" is only defined at the atomic positions, and so it only makes sense to speak of waves with wavevector k within the first Brillouin zone. For any wave with k outside the first BZ we can always find another one with $k \in 1^{st}$BZ which gives exactly the same atomic displacements. This is not a quantum mechanical effect - it is simply because a crystal

tions. Thermal conductivities of real materials are affected by such factors, yet they remain finite even when one makes bigger and better specimens.

[17] We noted earlier that these inelastic peaks are mysteriously broadened. Now we can explain why - a phonon with energy $\hbar\omega$ is not a true eigenstate of the crystal Hamiltonian and so decays with time. This "localisation in time" implies delocalisation in energy, i.e. its energy is not precise.

[18] Recall that metals conduct heat much better than non-metals and so the "free" electrons which allow electric currents to flow must also be the primary means of thermal conduction in metals. We might suppose that the electron scattering mechanisms (which we have yet to specify) giving rise to electrical resistance must also contribute to thermal resistance. Nonetheless phonon-phonon scattering occurs in metals and non-metals alike.

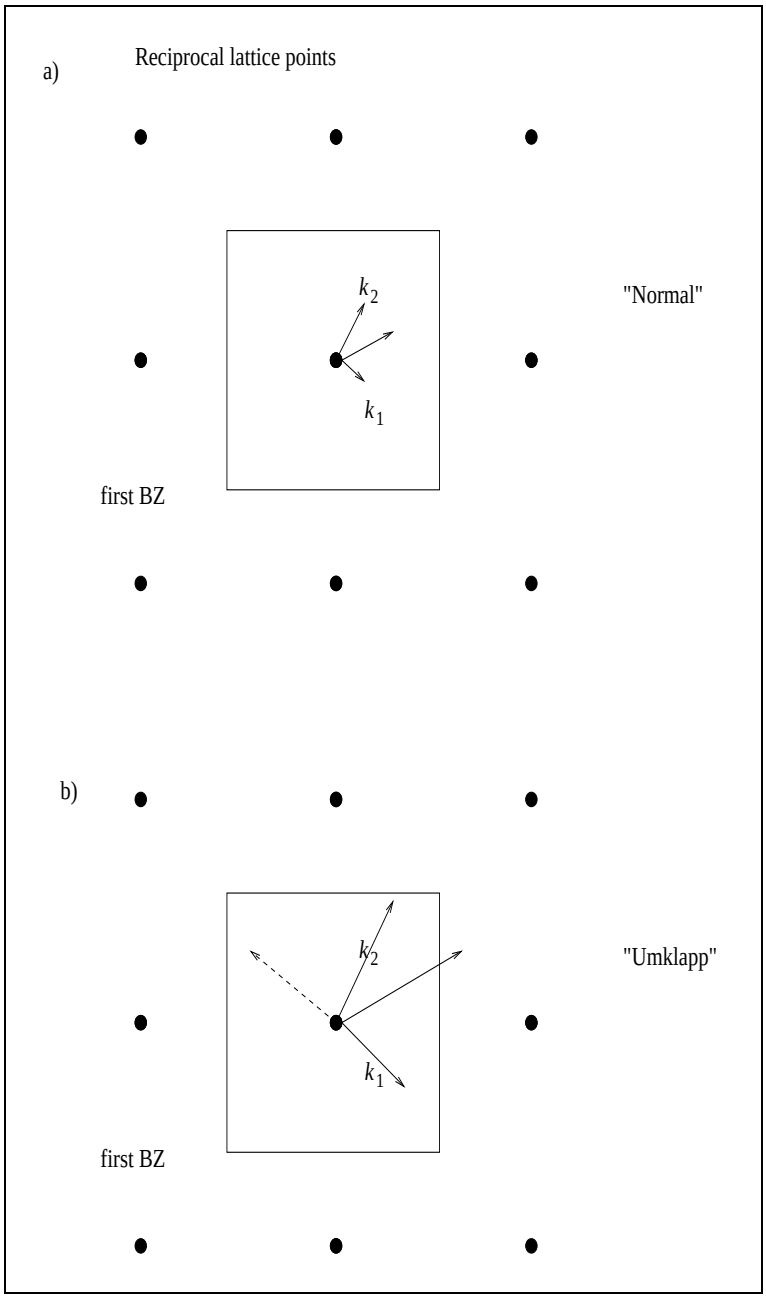

Figure 3.14: Combination of phonon wavevectors during phonon-phonon collisions in a simple cubic lattice. a) shows a normal process while b) shows an Umklapp process. The first Brillouin zone is also shown.

is made up of discrete atoms rather than a continuum.

Applying periodic boundary conditions we found that a system of N atoms supports wave-like modes of vibration with N distinct wavevectors. Solving the Newtonian equations of motion we obtained the corresponding allowed angular frequencies ω. We recovered the appropriate continuum limit as $|k| \ll k_{BZ} = \pi/a$ and found that ω tails off to a maximum (cut-off) frequency as k approaches the Brillouin zone boundary. We noted that the group velocity[19] at this point is zero, and we interpreted this in terms of standing waves caused by the Bragg diffraction condition.[20]

For the diatomic linear chain a new effect emerged. With N units cells we still have N allowed wavevectors which we can again choose to lie in the first BZ, but this time there are two solutions to the equations of motion for each k, giving two branches. In the lower branch the atoms within each unit cell tend to move in phase with each other, while in the upper one they tend to move out of phase. The lower branch was designated "acoustic" since it behaves as a sound wave as $k \to 0$, while the upper one was termed "optical" since the out of phase oscillation of chemically distinct atoms gives rise to an oscillating electric dipole which strongly interacts with electromagnetic waves.

The new feature of three dimensional systems was *polarization*, i.e. there are waves with the displacement perpendicular to as well along the direction of travel. For N unit cells and periodic boundary conditions we still have N allowed wavevectors. If there are p atoms in the basis, then there are $3p$ branches. In real solids one expects that one needs to account for more than nearest neighbour interactions. We dodged this issue by noting that so long as only harmonic forces are present it is always possible to convert the equations of motion of the system to give a set of equations describing independent harmonic oscillators.[21]

To explain the heat capacity of solids, the main motivation for this chapter, a quantum mechanical treatment of the harmonic modes of vibration was required. We quoted the basic result from Quantum Physics 3 that the allowed energies of the quantum harmonic oscillator have the form $(j+1/2)\hbar\omega$. When $k_b T \gg \hbar\omega$ the energy levels are so finely spaced that it doesn't really matter that they are discrete, but when $k_b T \ll \hbar\omega$ there is a profound change in behaviour. It is no longer possible to thermally excite an oscillator (with significant probability) and the heat capacity vanishes. We saw that the Einstein model, which builds in this basic idea, is a bit too crude to reproduce the detailed behaviour of C_V as $T \to 0$. The Debye model accounts for a distribution (albeit an approximate one) of allowed vibrational modes and reproduces the experimentally observed T^3 behaviour in the low T limit.

We noted some similarities between the Debye model for lattice vibrations and the qunatum theory of the free electron gas. In each case we are considering waves in a crystal subject to periodic boundary conditions. But we should be clear about a few things.

• A lattice vibration involves the collective motion of all the atoms in the crystal

[19] A *wavepacket* made up of a number of waves with wavevectors distributed about a particular k transports energy at the group velocity at this wavevector.

[20] In three dimensions the group velocity vanishes as $\underline{k}$ approaches the BZ boundary in a direction perpendicular to it.

[21] A more sophisticated treatment of this issue would point out that the appropriate canonical transformation is provided by expressing the atomic displacements as Fourier transforms. Eq. 3.10 is a simplification of the general case.

while the free electron wavefunctions considered in chapter 1 each described a single electron.
• The allowed $\underline{k}$ vectors are spaced by $2\pi/L$ in each dimension for both electron and lattice waves, but k_D sets an upper bound on the *allowed* wavevectors of lattice waves. For electron waves there is no cut-off in the allowed wavevectors.
• The Fermi wavevector k_F sets an upper bound on the *occupation* of the allowed states when $T = 0$ to account for the Pauli principle. There is no such restriction of the occupation of the allowed states of the quantum harmonic oscillators.

We had hoped to learn more about the origins of electrical resistance of metals, i.e. the agents that scatter conduction electrons. Could phonons play a role? The answer is "yes", but first we need to understand more about the interaction of electrons with the static lattice of a crystal. We have seen no hint of it yet but we may see at the end of the course that phonons can play a spectacular and totally unexpected role in determining the electrical conductivity of some solids by making rather subtle interventions in the interactions between electrons.

We should also note that we in Sec. 3.6 we briefly considered phonon-photon interactions. Extending our semi-classical treatment of diffraction we stumbled across an important conservation law: *the conservation of crystal momentum.* "Crystal momentum" is simply the name given to the quantity $\hbar\underline{k}$, where $\underline{k}$ is the wavevector of an allowed wave in a crystal, but you should be able to see some resemblance to true momentum. Don't worry about the origins of crystal momentum too much, but know that it is conserved to within $\hbar\underline{K}$. Nor should you worry about the details of neutron diffraction. It is enough to know why neutrons are suitable for both diffraction (elastic scattering) and phonon spectroscopy (inelastic scattering), while photons (x-ray and visible) are less so.

Having discussed the structures of crystals and how the atoms within them can move we should have the basis for understanding most of the mechanical properties of solids. The simplest mechanical experiment, however, still gives us some surprises. Imagine stretching a solid. All solids obey Hooke's law up to a point, but then metals tend to "yield" while non-metals tend to break. This makes some sense in terms of our discussion of chemical bonding. As we stress a metal there comes a point where planes of atoms start to slide past each other. Metals are not so fussy about the precise location of their neighbours, just that there be as many of them as possible, so distorting their structure does not cost a great deal of energy. Knowledge of some basic parameters, such as the cohesive energy of a metal, allows its *shear strength* to be estimated. Even sophisticated calculations get this wrong by four to five orders of magnitude! Planes of atoms slip past each other extraordinarily easily in metals. Covalent solids on the other hand rely on strong directional bonds for their cohesion. If we distort a covalent solid beyond a critical amount the atoms wind up in the wrong geometry for the bonds to work and the solids fractures. But this too happens in real samples far more easily than we can explain. There is an important concept (largely) missing from our discussion of crystal dynamics in this chapter.

Chapter 4

Introduction to band theory and the electronic structure of solids

We started the course by pondering the enormous diversity of condensed matter, with particular reference to electrical conductivity. Even among perfect crystals of elemental solids the range is quite extraordinary. In this chapter we will find the reason for this.

A discussion of the behaviour of electrons in solids can be conveniently conducted in two parts, as we saw in Chapter 1. Firstly we must consider solving the Schrödinger equation to obtain the stationary states of the system. While we solved the Schrödinger equation for free electrons in Chapter 1, a more realistic and general approach is now needed. The second ingredient is the inclusion of scattering effects.[1]

The term *electronic structure* refers to the energy levels of electrons and their distributions in space and momentum. Their characterization, calculation and experimental investigation are central goals of condensed matter physics, materials science and chemistry. We start this chapter with a discussion of some quite fundamental points:

1. We will assume that only the outermost electrons of atoms are of interest when a solid is formed. The inner or "core" electrons remain tightly bound to the nucleus, which together constitute the ion core. While only the outer or "valence" electrons participate in chemical bonding, the core electrons play an indirect but spectacular role. This can be seen by comparing diamond, silicon, germanium, tin and lead. These elements are all in Group 4 of the periodic table and their atoms each have an ion core with charge 4+ and four valence electrons. Does it follow that these solids have the same properties? No. Diamond is an extremely hard, covalently bonded insulator, silicon and germanium are covalent but conduct electricity (though poorly) at room temperature, tin exists as either a metal or a semiconductor, while lead is a very soft metal. We

[1] Of course we would like to put absolutely everything into the Schrödinger equation, but it is hard enough to solve even for a perfect (non-vibrating) crystal. The things we neglect at the Schrödinger equation level and have to be cobbled on afterwards.

will see that these profound differences can be attributed in large part to the influence of the core electrons.

2. We will assume that the ion cores in a crystal are arranged with perfect periodicity, and are therefore stationary. This sits rather uncomfortably with the previous chapter. By the same token it seems a bit strange that one can study lattice dynamics without the slightest regard for the role of electrons. This dichotomy in condensed matter physics between those properties that can be explained by atomic vibrations and those attributed to electrons requires some justification. Electronic and core motions are separable[2] to a good approximation because the mass of the electron m is extremely small compared to ion core masses M (m/M is between 10^{-4} and 10^{-5} in most cases). It follows that electrons move extremely quickly compared to core motions and so they very rapidly adapt to changes in atomic position. The assumption that the electrons are always in the ground state appropriate to the instantaneous ionic configuration (regardless of whether the ions are moving) is called the *Born-Oppenheimer approximation* or the *adiabatic approximation*. The separation is not perfect, as we shall see.

3. There remains the basic question raised in Chapter 1 regarding the extent to which the electrons are independent of *each other*. Certainly it is very convenient to regard each electron as having its own wavefunction and as being in a definite quantum state. However, the idea of describing an interacting system in terms of individual wavefunctions is incorrect. Formally, when the Hamiltonian of a system contains interactions between particles the Schrödinger equation does not separate and the wavefunction of the system cannot be built up from single particle wavefunctions. Likewise, one cannot hope to learn much about the game of chess simply from a knowledge of how the individual pieces are allowed to move on an empty board.

Despite these quite discouraging remarks, the naive idea that each electron should have its own wavefunction and eigenenergy turns out to be extremely useful. This is a subtle but very important point: a "one electron" approach does not necessarily require there be *no interactions at all* between electrons. Rather it requires them to influence each other only through the establishment of an average potential.[3] In this chapter we will assume that an electron in a solid feels a periodic potential (due to all the other particles in the system) which depends solely on its position. Solving the Schrödinger equation for an electron in this potential will give us a set of energy levels which we will successively fill up in accordance with the Pauli principle. The majority (though not all) of experimental results and observed phenomena can be explained in these terms.

4. In the previous chapter we saw that room temperature can usually be considered "high" when considering lattice vibrations. On the other hand we saw in Chapter 1 that typical electronic energies are rather large compared to normal thermal energies. We may therefore expect that all temperatures of interest are

[2]Formally we can express the wavefunction of a solid as the product of an electronic part and a part describing how atomic motions, as we shall see later.

[3]Perhaps we should have used the analogy of a chess piece participating in a game in which all other pieces are spread over their average positions.

"low" as far as the electrons in solids are concerned. This is substantially true, with one very important exception.

These "introductory" remarks are not minor preliminaries to be quickly dismissed, but are some of the fundamental issues in condensed matter physics.

4.1 Electrons in condensed matter

The Schrödinger equation for a crystal is[4]

$$H\Psi = \sum_i \left\{ -\frac{\hbar^2}{2m}\nabla_i^2 - \sum_R \frac{Ze^2}{4\pi\varepsilon_0 \left|\underline{r}_i - \underline{R}\right|} \right\} \Psi + \frac{1}{2} \sum_{i,j\ (i\neq j)} \frac{e^2}{4\pi\varepsilon_0 \left|\underline{r}_i - \underline{r}_j\right|} \Psi = E\Psi \tag{4.1}$$

where the first sum is over all the electrons in the solid, the R sum is over Bravais lattice points,[5] and the third sum is over all pairs of electrons in the solid. In writing down this equation we are assuming that we know where the atoms are located and we are assuming that they are stationary. The wavefunction Ψ describes only the electrons in the solid, but it includes all of them. Clearly we have a *many body problem* which we cannot solve. To make progress we make the *one electron approximation* (OEA) and assume that the wavefunction for the whole electronic system can be expressed as a product of wavefunctions which each describe only one electron. If the two potential energy terms in Eq. 4.1 can be written

$$-\sum_{i,R} \frac{Ze^2}{4\pi\varepsilon_0 \left|\underline{r}_i - \underline{R}\right|} + \frac{1}{2} \sum_{i,j\ (i\neq j)} \frac{e^2}{4\pi\varepsilon_0 \left|\underline{r}_i - \underline{r}_j\right|} \approx \sum_i U_i(\underline{r}_i). \tag{4.2}$$

This means that each electron i is assumed to move under the influence of a potential $U(\underline{r}_i)$. The important thing here is that the potential for electron i depends only on the position of that electron and not explicitly on the positions of the others. This means that the electronic motions are *uncorrelated* and we should be able to use the old mathematical trick of "separation of variables". The true electronic wavefunction is an unknown function of the coordinates of *all* the electrons which we could write as $\Psi(\underline{r}_1, \underline{r}_2, \underline{r}_3, \cdots)$, but if the electrons are uncorrelated then we can replace Ψ with the product $\psi_1(\underline{r}_1)\,\psi_2(\underline{r}_2)\,\psi_3(\underline{r}_3)\cdots$, where each ψ describes only one electron. In this way the many body Schrödinger equation separates into a number of one electron Schrödinger equations with the form:

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\underline{r}) + U(\underline{r})\psi(\underline{r}) = \epsilon\psi(\underline{r}). \tag{4.3}$$

Notice that we have not neglected the electron-electron and electron-nucleus interactions, rather they enter Eq. 4.3 in some kind of approximate way. (Note that if the electrons do not interact then the OEA is exact and the total electronic energy E is the sum of the single electron eigenvalues ϵ_i, but in general this is not true.)

[4] We are ignoring lattice dynamics so Ψ describes only the electronic degrees of freedom.

[5] We assume a single atom basis for notational brevity.

Before we get too carried away with this simplification we should remember that it really is an approximation. Take a hydrogen molecule, for example. If at a particular instant the first electron happens to be on atom A, then we would strongly suspect that the second electron will be on atom B. But in the scheme outlined above we have stated that the whereabouts of electron two are determined solely by the potential $U(\underline{r}_2)$ which is independent of the position of electron one at the moment in question. In the one-electron approximation it is equally likely for the electrons to be on the same atom as it is for them to be on different atoms! Electron motions really are correlated - they must be because electrons repel each other and therefore move in such a way as to avoid each other.

But we need not despair on account of the rather extreme example referred to above. The trick is to construct the potential U so that it accounts for electron-electron correlations as best we can. This task is rather hard since the ingredients of U are the electron-ion interactions and the electron-electron interactions,[6] and to estimate the repulsion any particular electron experiences from all the other electrons would require us to know the wavefunctions of all the others, which one does not know at the outset. The way to proceed is to *guess* the ψ_i at the start, construct U, solve Eq. 4.2 to get new estimates of the ψ_i, then check for consistency with the old wavefunctions. We must go round this self-consistency loop until the wavefunctions we get out are consistent with the one we use to compute the electron-electron interactions.

The *self-consistent field method* is shown schematically in Fig. 4.1 and we must think about its implementation. Let's postpone that for now and simply assume that we somehow know the self-consistent potential U, concentrating on how to solve Eq. 4.3 for electrons in a crystal and the properties of the wavefunctions we obtain.

4.2 The empty lattice approximation

It is instructive to first consider a crystal where the potential U is so weak it is vanishingly small. This so-called *empty lattice approximation* is clearly just a form of free electron model, so what's the point of it? It serves to introduce the way we display electron states in crystals.

We saw in Chapter 1 that the electron energies and wavefunctions for the free electron model are $\epsilon = \hbar^2 k^2/2m$ and $\psi(\underline{r}) \propto e^{i\underline{k}.\underline{r}}$ respectively. We may ask which of these states is occupied (at $T = 0$). In Section 1.2 we showed that the density of allowed electron wavevectors in a crystal is $W/(2\pi)^3$, where W is the volume of the entire crystal. Since each $\underline{k}$-state can accommodate 2 electrons (because of spin degeneracy) the first Brillouin zone contains sufficient $\underline{k}$ points to label the wavefunctions of $2W/V$ electrons,[7] where V is the volume in real space of the unit cell. We sometimes simply say that "each BZ can accommodate $2W/V$ electrons". Higher order BZ's each have the same volume as the first, and so each can accommodate $2N$ electrons. $2W/V$ is equal to $2N$, twice the number of primitive cells in the crystal, and so if there are two valence

[6] It is essential to appreciate that neglecting electron-electron *correlations* is not the same as neglecting electron-electron *interactions*. We include the later approximtely by including an electron-electron repulsion contribution to U.

[7] Recall that the volume of the primitive unit cell of a reciprocal lattice is equal to $(2\pi)^3/V$.

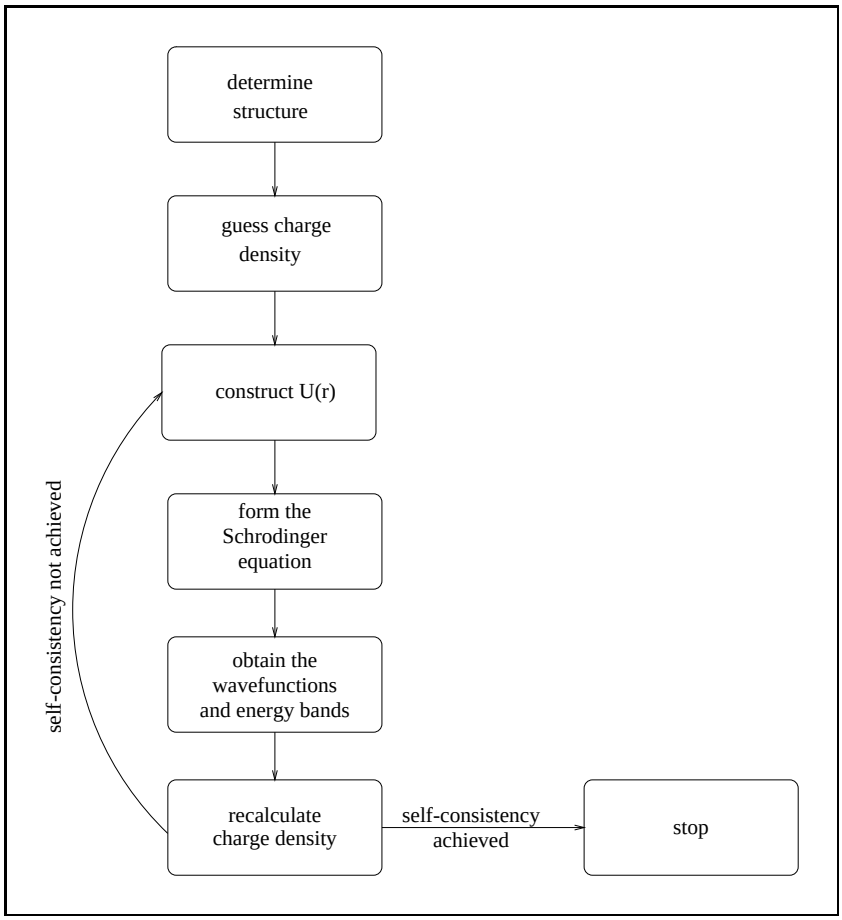

Figure 4.1: The steps in calculating the electronic states of a solid.

electrons per primitive cell then there are enough to fill the first Brillouin zone.

The 1D free electron dispersion relation is shown as the solid line in Fig. 4.2. If there is one electron per unit cell then the first BZ will be half full, assuming the states fill up starting with the lowest energy first. For two electrons per cell the first BZ is full, and so on. This representation is called the *extended zone scheme*. For compactness it is often convenient to map all states into the first BZ by translation by the appropriate reciprocal lattice vector. In this *reduced zone scheme*, also shown in Fig. 4.2, the dispersion relation has the form

$$\epsilon(\underline{k}) = \frac{\hbar^2(\underline{k} - \underline{K})^2}{2m} \tag{4.4}$$

where $\underline{K}$ is a reciprocal lattice vector. When $\underline{k}$ is confined to the first BZ, $\epsilon(\underline{k})$ becomes a multi-valued function and it is therefore necessary to label each portion of $\epsilon(\underline{k})$ or *band* with a different *band index*: $\epsilon_n(\underline{k})$.[8] Since in the extended zone scheme each BZ can hold $2N$ electrons, each band in the reduced zone scheme also holds up to $2N$ electrons.

[8]We can assign band indices according to which reciprocal lattice vector is required to translate the band into the first BZ, but this is not the only way.

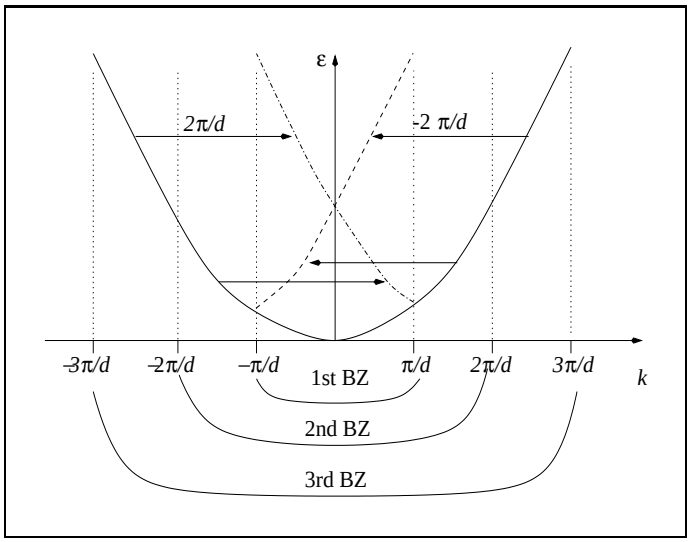

Figure 4.2: Free electron energy levels for a 1D chain with atomic spacing d. The solid curve shows the extended zone representation. Also shown is the reduced zone scheme constructed by translating all portions of the $\epsilon(k)$ curve by a reciprocal lattice vector so as to confine $\underline{k}$ to the first Brillouin zone. We call the dashed and dot-dashed lines the second and third *bands*.

The representation of the *electron energy bands* in the reduced zone scheme, as shown in Fig. 4.2, looks reminiscent of the *phonon frequency branches* from Chapter 3, but we should again stress the important differences.
(i) Notice that there is no cut-off energy for the electronic states. There are an infinite number of energy bands.
(ii) We showed that a lattice vibration with k outside the first BZ gives exactly the same atomic displacements as a mode with wavevector within the first BZ. We have not made the same claim for electrons. Our use of the reduced zone scheme must be viewed solely as a convenience, at this stage at least.
(iii) The occupation of the vibrational modes is determined by the Bose-Einstein function, while the occupation of the electronic states follows Fermi-Dirac statistics.

The filling of the energy bands in more than one dimension can lead to some surprises. Energy bands and the Fermi sphere in the first two Brillouin zones of a 2D square lattice with two valence electrons per atom are shown in Fig. 4.3 in both the extended (on the left) and reduced (on the right) schemes. The first thing to note is that the radius of the Fermi sphere has roughly the same dimensions as the first Brillouin zone. In fact they have the same area (i.e. contain the same number of states) since for the particular case shown there are two valence electrons per primitive cell. But because their shapes are different the first BZ is not completely filled. It *could* accommodate two electrons per cell, but it doesn't since there are parts of the second zone with lower energy. This figure is extremely important - we will see why in the next section.

For a 3D system, the task of displaying $\epsilon(\underline{k})$ vs $\underline{k}$, called the *band structure* of a crystal, is difficult since there are four independent components (three for the components of $\underline{k}$ and one for energy). This is usually accomplished by plotting

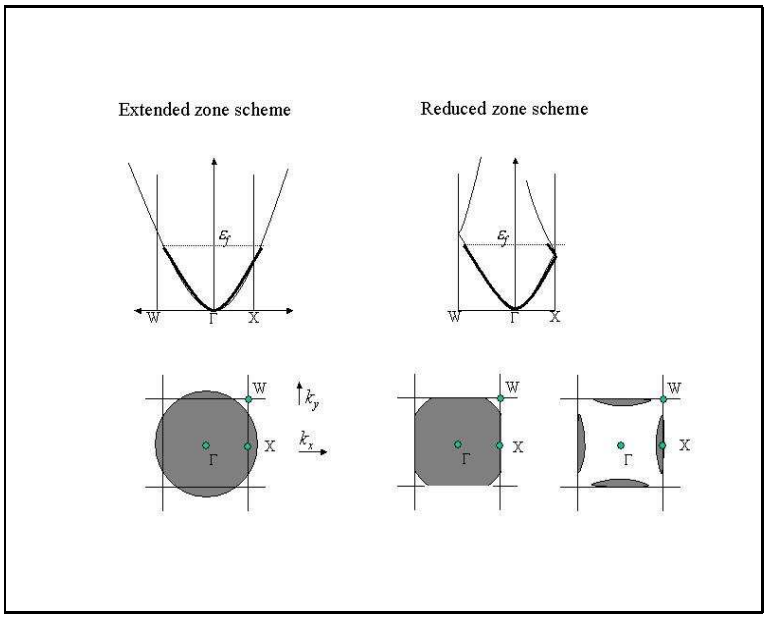

Figure 4.3: The first two Brillouin zones of the 2D empty square lattice with sufficient electrons to fill precisely one zone.

$\epsilon(\underline{k})$ vs $\underline{k}$ in the reduced zone scheme for $\underline{k}$ along just a few directions across the first BZ.[9] The complexity of this representation can be surprisingly great, even for the empty lattice, as shown in Fig. 4.4 for the FCC structure. As one should expect, many of these bands are degenerate due to the high symmetry of the chosen directions and the free electron states.[10]

4.3 Nearly free electron model

4.3.1 The allowed states

Now let's consider what happens when the crystal potential U is non-zero, but weak. It is natural to expect that the free electron picture of the empty lattice approximation will have to be modified slightly. We can anticipate the physics of the *nearly free electron model* (NFEM) by wondering what would happen to a free electron propagating through a crystal with a weak crystal potential. In fact we have already considered this situation (i.e. a plane wave interacting weakly with a crystal) in another guise - x-ray diffraction. For brevity we consider the 1D case (with spacing a) for which the reciprocal lattice vectors are $K = 2j\pi/a$, where j is any (positive or negative) integer, and the Bragg condition is just $k = j\pi/a$. Let's just consider first order diffraction for now. Since $\theta = \pi/2$ for a 1D system (see Fig. 4.5), the incident and reflected waves combine to form standing waves.

[9] To be more precise, we usually plot the band structure between symmetry points in an irreducible "wedge" of the first BZ.

[10] When the crystal potential is non-zero some of these degeneracies are lifted.

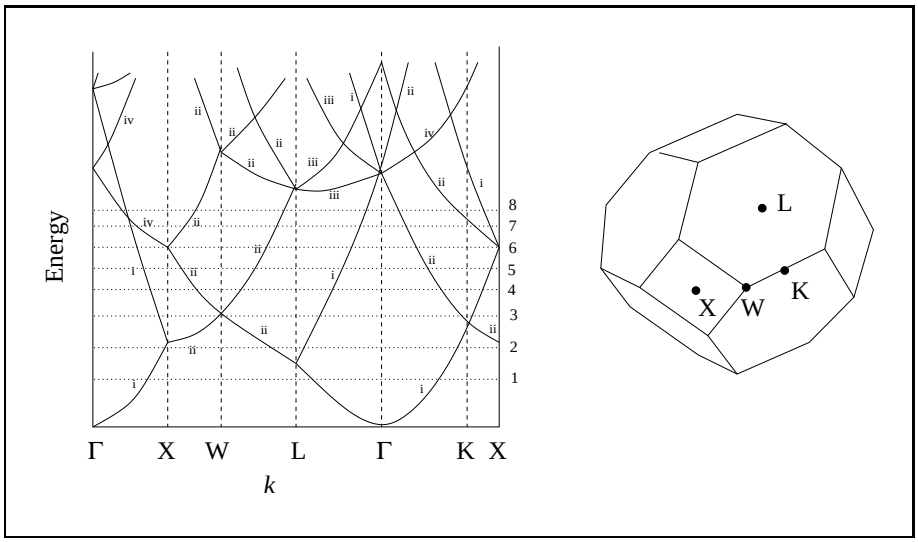

Figure 4.4: Free electron energy levels for the FCC Bravais lattice along certain directions in the Brillouin zone using the reduced zone scheme. The horizontal dotted lines denote the Fermi energies for the indicated numbers of electrons per primitive unit cell. The roman numerals indicate the degeneracy of each band. The BZ is shown on the right. The Γ point is at the centre of the BZ.

Different standing waves can be formed depending on the relative phase of the component waves. The two extremes are shown in Fig. 4.6 and given by

$$\psi_1(x) = e^{i\pi x/a} + e^{-i\pi x/a} = 2\cos(\pi x/a)$$
$$\psi_2(x) = e^{i\pi x/a} - e^{-i\pi x/a} = 2i\sin(\pi x/a). \tag{4.5}$$

In the free electron model the electron density for each level $\psi^*(x)\psi(x)$ is independent of position, since there is nothing to make it otherwise. Once the electrons can "feel" the lattice potential things are different. ψ_1 has most of its density at the ion sites, while ψ_2 has its density concentrated in between them, as shown in Fig. 4.6. The electron-nucleus interaction is attractive and so ψ_1 has the lower energy. Can we quantify this?

Since we are postulating a weak crystal potential we can be optimistic about correcting the energies of the free electron model by recourse to first order perturbation theory, for which the energy correction is

$$\Delta\epsilon = \frac{\langle\psi|U|\psi\rangle}{\langle\psi|\psi\rangle} \tag{4.6}$$

where the ψ are the unperturbed wavefunctions (i.e. free electron plane waves). The potential $U(x)$ is periodic with period a and so can be expressed as a Fourier series (Physical Mathemtics from last year?):

$$U(x) = \sum_K U(K)\, e^{iKx}. \tag{4.7}$$

It is instructive to use a rather short Fourier expansion:

$$U(x) \approx U_0 + 2U_1\cos(2\pi x/a) \tag{4.8}$$

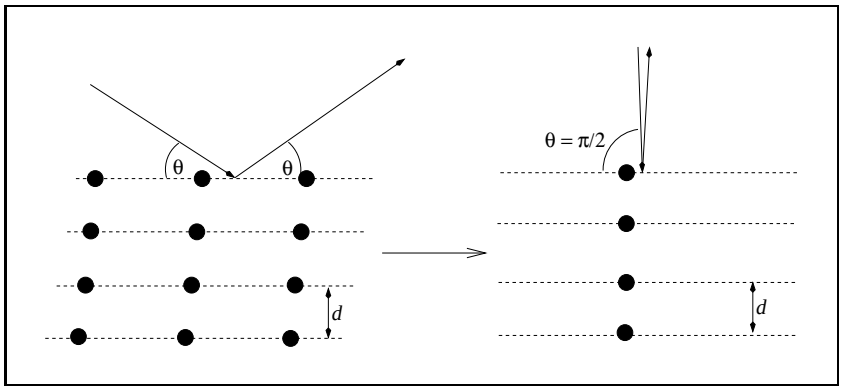

Figure 4.5: Bragg reflection in 1D. The Bragg "planes" become single sites while the incident angle must be $\pi/2$.

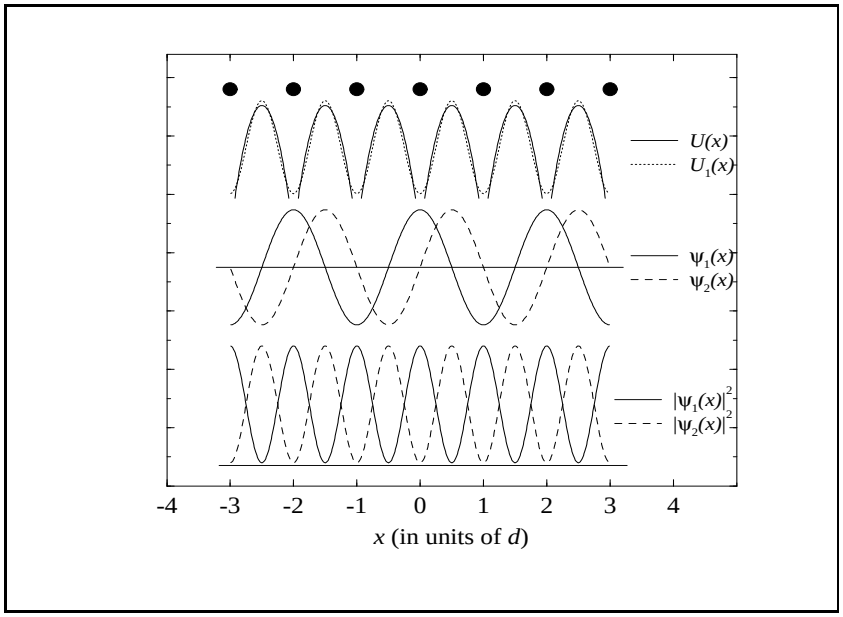

Figure 4.6: Schematic representation of the potential $U(x)$ for a 1D chain of atoms and its first Fourier component (a cosine function). The standing waves ψ_1 and ψ_2 produced by Bragg reflection are shown, as well as the corresponding electron densities $|\psi_1|^2$ and $|\psi_2|^2$.

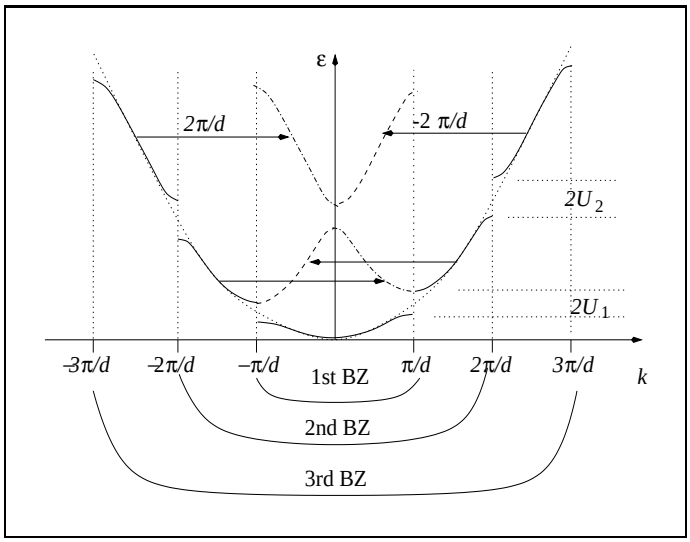

Figure 4.7: Schematic representation of the energy levels for the nearly free electron model in the extended and reduced zone schemes. (Note d should be a for consistency with text.)

where U_0 and $U_1 = U(2\pi/a) = U(-2\pi/a)$ are negative constants. It would appear to be a simple matter to combine Eq. 4.6 and 4.8 but the result is strange: we get $\Delta\epsilon = 0$, while second and higher order corrections are infinite. The problem here is a technical one. The wavefunctions $\exp(\pm i\pi x/a)$ have the same energy and so we must use *degenerate perturbation theory*. The rule here is that one must start with unperturbed eigenfunctions which are also eigenfunctions of the perturbation. In the present case, the linear combinations given in Eq. 4.5 make suitable zero-order wavefunctions, giving first order energy corrections $U_0 \pm U_1$.

It can be shown that the $\epsilon(k)$ curve approaches the lower energy solution as k approaches the first BZ boundary from the interior of the zone (i.e. as $k \to \pi/a$ for the first energy band in the reduced zone scheme). Approaching the first BZ boundary from outside (i.e. as $k \to \pi/a$ for the second energy band in the reduced zone scheme), $\epsilon(k)$ converges to the higher energy solution. In other words, there is a range of energies, known as a *band gap* for which there are no allowed electronic states.

Since ψ_1 and ψ_2 are standing waves, the group velocity $(1/\hbar)d\epsilon/dk$ is zero at Brillouin zone boundaries. Non-zero higher order Fourier components also give rise to discontinuities in the dispersion relation when k lies on higher order Brillouin zone boundaries. These results are summarized in Fig. 4.7.

4.3.2 Occupation of the states

As before we use the Pauli principle i.e. fill the lowest energy states first, putting only one electron per state. Remember that strictly speaking this is only correct for $T = 0$, but since $k_b T << \epsilon_f$ for any relevant T, then the $T = 0$ limit seems appropriate.

In one dimension things are simple. For one electron per unit cell the first

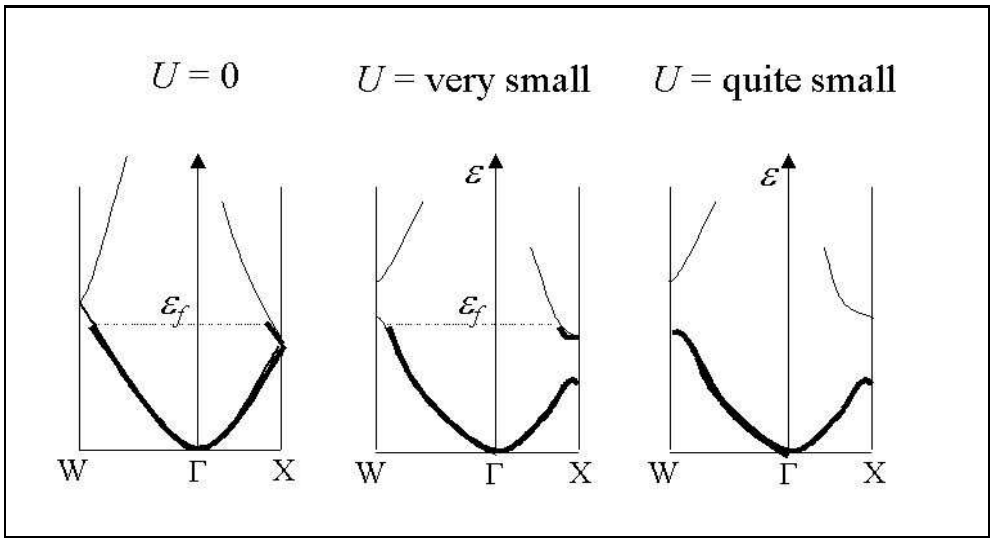

Figure 4.8: Filling of the energy bands for the 2D square lattice when there are two electrons per unit cell. Three different strength of crystal potential U are considered.

band is half filled, for two electrons per unit cell the first band is filled, for three electrons per unit cell the second band is half filled, etc. In two and three dimensions things are not so simple. Certainly for an odd number of electrons per unit cell we will have unfilled bands. But for an even number of electrons per unit cell we can get qualitatively different behaviour depending on the size of U, as illustrated in Fig. 4.8. If the crystal potential is very weak (and hence the band gaps at the BZ boundaries are very small) then the "empty lattice" picture is not significantly modified. However, if U is sufficiently strong it can be seen that the energy bands can become completely filled or completely empty. The crucial issue is the size of the gaps. Now let's discuss the significance of Fig. 4.8.

4.3.3 Metals and Insulators

There is no greater concept in solid state physics than that of the band gap. It is the basis of our understanding of the distinction between metals and insulators (and semiconductors).

Metals

Cast your mind back to Chapter 1 where we started to think about the conduction of electricity in the free electron gas. We had a quantum mechanical description of the electronic states but a classical description of electron dynamics in the presence of an external force. Electrical conduction was explained in terms of a balance between the acceleration of the free electrons by an external DC electrical field and the damping effect of some kind of scattering mechanism(s). Implicit in this picture is the assumption that the conduction electrons can have their velocities increased. In the free electron gas we imagine the Fermi

sphere tending to shift so that it is no longer centred on the origin, as shown in Fig. 4.9. (Of course electron scattering limits this shift and a dynamic steady state is set up.) Since the energy bands in the free electron gas are continuous, it is certain that for an infinitesimal shift of the Fermi sphere there will be empty states which are only an infinitesimal energy above the Fermi energy. Thus the energy cost of producing an electron distribution which corresponds to a current-carrying state is infinitesimal. We should also observe at this point that we are confidently associating the wavevector (and hence momentum) of a free electron with its velocity, give or take some constants. This seems totally sensible but we will have more to say on that soon.

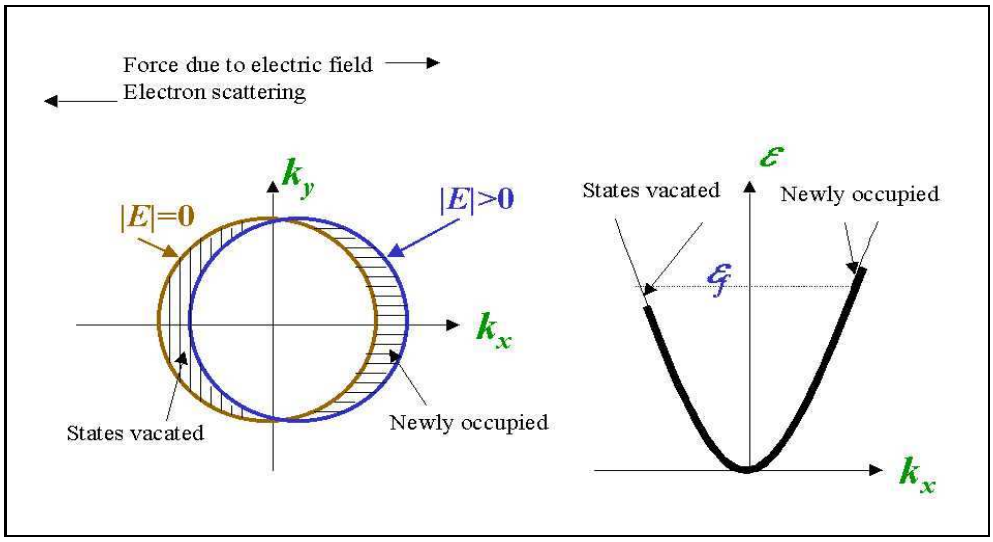

Figure 4.9: Shifting of the Fermi sphere of a free electron gas by an external potential.

The NFEM results may have also alerted you to an explanation of the existence of electrical insulators. Consider a crystal (for which the NFEM is valid) with one electron per unit cell. There are enough electrons to exactly half fill a single electron energy band (which means enough to occupy half of the $\underline{k}$ states in a band). Since energy bands are continuous, such a system **must** be a conductor since there are empty states an infinitesimal distance above the Fermi energy. In fact we can state with certainty that any crystal with an odd number of valence electrons per unit cell must be metallic since it must have at least one partially filled band.

In metals there will be a surface in k-space separating the occupied and unoccupied states of the partially filled bands. These are called *Fermi surfaces* and are a generalization of the Fermi sphere of the free electron gas. In the final Problem Sheet we will explore the Fermi surface within the NFEM. In particular we will see how Fig. 4.2 is modified by the crystal potential.

Semi-metals

There is the possibility of a metal having an exceedingly small Fermi surface (i.e. all the partially filled bands are either very nearly full or very nearly empty).

Such solids tend to conduct poorly and so are frequently called *semi-metals*. Here we see a criterion for *quantifying* electrical conductivity: The greater the Fermi surface, the greater the electrical conductivity.

Insulators

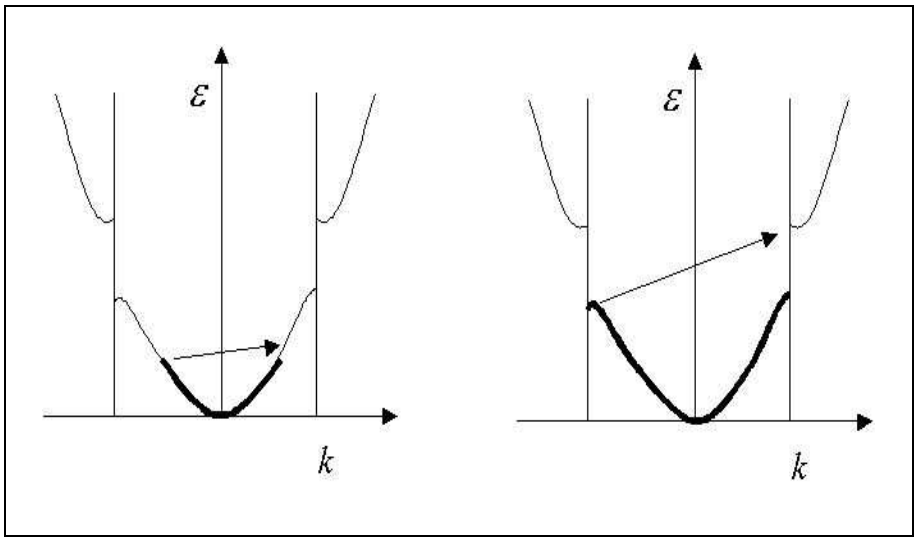

Figure 4.10: Energy bands for a one dimensional system. The system on the left has a single electron per unit cell. There is no energy gap between the highest occupied state and the lowest empty state so we have a conductor. For the case illustrated on the right there are two electrons per unit cell. There is a large energy cost preventing the formation of a current-carrying state.

Now consider the case of an even number of electrons per unit cell combined with a strong crystal potential. We have seen that these conditions lead to a complete absence of partially filled energy bands; each band is either filled or empty. There is a distinct energy gap between the highest occupied state and the lowest empty state, as illustrated in Fig. 4.10. By definition, such solids do not have a Fermi surface since they do not have partially filled bands. Although an extremely strong electric field may be capable of ripping an electron from a filled band and promoting it to a previously empty band, a process known as *dielectric breakdown*, electric currents cannot be produced by connecting a battery to a wire of the material. The energy required is far too great.

Semi-conductors

If the energy gap ϵ_g between the highest occupied and lowest unoccupied levels is of order 1 eV, then we find room temperature conductivities intermediate between those of typical metals and insulators. We usually refer to narrow-band-gap insulators as *semiconductors*. At room temperature $\epsilon_g/k_bT \sim 40$ and there is a non-negligible probability of thermal excitation of an electron from

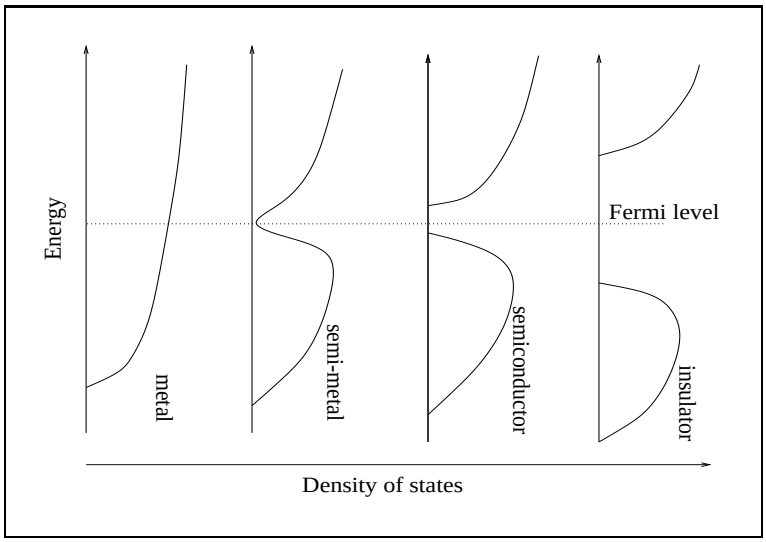

Figure 4.11: The density of states for a fictitious metal, semi-metal, semiconductor and insulator.

the highest occupied to the lowest unoccupied band.[11]

The density of energy levels

We have seen how electron energy bands are plotted. When we are only interested in electron energies and not their wavevectors, it is convenient to compress thr full band structure information into the form of a density of levels (or states) g, as we did for free electrons in Chapter 1 and for phonons in Chapter 3. Applying Born-Von Karman (periodic) boundary conditions to Eq. 4.11, it follows that the volume (in k-space) of each allowed k-point is $(2\pi)^3/W$, and hence $g^{\underline{k}} = W/(8\pi^3)$, just as in the free electron model. We can then equate $g^{\epsilon} d\epsilon$ with twice the number of allowed k states with energy in the range ϵ to $\epsilon + d\epsilon$.

The four classes of solids discussed above are easily distinguished according to their density of states, as illustrated by Fig. 4.11. But note that the distinctions between metals and semi-metals and between semiconductors and insulators are not fundamental.

[11] Incidentally, many insulators have conductivities which, although exceedingly small, are still too large to be explained by thermal promotion across the band gap.

4.4 Electrons in a periodic potential

The NFEM provides an explanation for the distinction between metals and non-metals, but provokes further questions:

• The electronic states resemble a single plane wave in the free electron limit, and a superposition of two plane waves near the BZ boundaries in the NFEM. What is the general form of an electronic state in a crystal?

• Can we better justify use of the reduced zone scheme?

• Why is the crystal potential weak for certain crystals?

We tackle the first of these in the next subsection.

4.4.1 Bloch's theorem

Consider a one dimensional chain of atoms with positions given by $R = ja$ where j can be any integer. Since every atom in the chain is identical one expects some kind of periodicty to emerge in the electron wavefunctions of the system. Let's start by considering the value a particular wavefunction takes at the origin $R = 0$, which we will denote $\psi(0) = z$. It is tempting to suppose that the periodicity of the system requires $\psi(a) = z$, but this is not the case. Certainly all physical properties must be periodic, but ψ is not a physical quantity. However the electron density ($|\psi|^2 = \psi\psi^*$) is, and so this must be periodic implying $\psi(0)$ and $\psi(a)$ can differ only by a phase factor: $\psi(a) = z\,e^{i\theta}$, and in general

$$\psi(R) = z\,e^{i\theta R/a}. \tag{4.9}$$

It appears that θ/a play the role of wavevector here, and we can write

$$\psi(R) = \text{const.} \times e^{ik.R}. \tag{4.10}$$

This is all fine if we are only interested in the values of ψ at the atomic positions R. This is the case for lattice vibrations which we studied in §3. In such cases we can readily justify the reduced zone scheme since adding a reciprocal lattice vector $K = (2\pi/a) \times$ integer onto k in the previous equation does not change the value of $\psi(R)$. Thus wavefunctions that are only defined at the positions R are periodic in k-space.

Electron wavefunctions are continuous functions in real space and this complicates things a little. The key is to notice that the argument given above can be repeated for positions $r = h + ja$, where h is some constant. As h varies, z must also vary, *but with periodicity* a. We have more or less proved[12] the one-dimensional version of Bloch's theorem:

The eigenstates of the one electron Schrödinger equation for a periodic potential have the form of a plane wave times a function which has the periodicity of the Bravais lattice

$$\psi_{\underline{k}}(\underline{r}) = e^{i\underline{k}.\underline{r}}\,u_{\underline{k}}(\underline{r}) \tag{4.11}$$

[12] We'll do a more respectable mathematical in a problem sheet.

where $u_{\underline{k}}(\underline{r}+\underline{R}) = u_{\underline{k}}(\underline{r})$ *for any Bravais lattice vector* $\underline{R}$.

These *Bloch waves* are planes waves modulated by some function which has the periodicity of the lattice.

4.4.2 Properties of Bloch functions

Crystal momentum

Bloch's theorem introduces a wavevector $\underline{k}$ which turns out to play the same role for electrons in a periodic potential as the free electron wavevector plays in the Sommerfeld model. Note that in the free electron gas $\hbar\underline{k}$ is the electron momentum $\underline{p}$, but this is not so for Bloch electrons. Indeed, the eigenstates of a periodic potential are not momentum eigenstates (i.e. momentum is not a constant of the electron motion). Nonetheless, $\hbar\underline{k}$ is a natural generalization of momentum for a periodic potential and to reinforce the point it is usually referred to as the *crystal momentum*.

Electron velocity

The *average* velocity of a Bloch electron is given by $\langle \underline{v} \rangle = \nabla_{\underline{k}} \epsilon(\underline{k})/\hbar$.[13] This may seem surprising. It implies that electrons in solids can travel indefinitely, without any degradation of their (average) velocity. This is far removed from the Drude-Lorentz or Sommerfeld descriptions of electron velocities.

Real space periodicity

Eq. 4.11 implies

$$\psi_{\underline{k}}(\underline{r}+\underline{R}) = e^{i\underline{k}.\underline{R}}\,\psi_{\underline{k}}(\underline{r}). \tag{4.12}$$

Thus Bloch waves are *not* periodic in real space (although their density $\psi_{\underline{k}}(\underline{r})\psi^*_{\underline{k}}(\underline{r})$ is).

It can also be shown (see Problem Sheet 1) that Eq. 4.12 implies Eq. 4.11, and so the two statements are equivalent.

Boundary conditions

Applying Born-Von Karman (periodic) boundary conditions to Eq. 4.12, it follows that the volume (in k-space) of each allowed k-point is $(2\pi)^3/W$, where W is the volume of the entire crystal, just as in the free electron model.

Plane wave expansion

The function $u_{\underline{k}}(\underline{r})$ in Eq. 4.11 has the periodicity of the direct lattice and so can be written as a Fourier series (i.e. as a sum over reciprocal lattice vectors):

$$u_{\underline{k}}(\underline{r}) = \sum_{\underline{K}} B^{\underline{k}}_{\underline{K}}\, e^{i\underline{K}.\underline{r}} \tag{4.13}$$

[13] Compare this with the classical result for a free particle in 1D: $v = d\epsilon/dp$. ($\nabla_{\underline{k}}$ means grad in k-space.)

and so the plane wave expansion of a Bloch wave takes the form

$$\psi_{\underline{k}}(\underline{r}) = \sum_{\underline{K}} B^{\underline{k}}_{\underline{K}}\, e^{i(\underline{K}+\underline{k})\cdot\underline{r}}. \tag{4.14}$$

In other words, a Bloch wave with wavevector $\underline{k}$ is a superposition of plane waves with wavevectors $\underline{k}+\underline{K}$.

A periodic crytsal potential only mixes those planes wave states whose wavevectors differ by reciprocal lattice vectors.

This observation should make us a little more comfortable about the reduced zone scheme. In general a Bloch wave with wavevector $\underline{k}$ will not necessarily resemble the plane wave with wavevector $\underline{k}$, and so there is no compelling reason to associate it with any particular BZ.

4.4.3 k-space periodicity and the reduced zone scheme

Up to now we have justified the reduced zone scheme for electronic states solely on the basis of convenience, but Bloch's theorem allows a rigorous justification. It is often stated that Bloch waves are periodic in reciprocal space. Inspection of the free electron wavefunction $\psi_{\underline{k}} = e^{i\underline{k}.\underline{r}}$ (which is a perfectly good Bloch wave) shows that this is not quite right. The correct justification for the reduced zone scheme is slightly more subtle.

Let's take a Bloch wave $\psi_{\underline{k}_1}(\underline{r})$ with wavevector $\underline{k}_1$ *outside* the first zone. There must be a reciprocal lattice vector that we can add to $\underline{k}_1$ to get a new vector $\underline{k}_2$ which is *inside* the first zone. Now we define a new wavefunction, which we can decide to call $\phi_{\underline{k}_2}(\underline{r})$, by the expression

$$\phi_{\underline{k}_2}(\underline{r}) = e^{i\underline{k}_1\cdot\underline{r}} u_{\underline{k}_1}(\underline{r}) = e^{i\underline{k}_2\cdot\underline{r}} e^{-i\underline{K}_0\cdot\underline{r}} u_{\underline{k}_2+\underline{K}_0}(\underline{r}). \tag{4.15}$$

The new wavefunction has the form $e^{i\underline{k}_2\cdot\underline{r}}$ multiplied by a more complicated function which we can call $v_{\underline{k}_2}(\underline{r})$. This function has the periodicity of the lattice since

$$v_{\underline{k}_2}(\underline{r}+\underline{R}) = e^{-i\underline{K}_0\cdot(\underline{r}+\underline{R})} u_{\underline{k}_2+\underline{K}_0}(\underline{r}+\underline{R}) = e^{-i\underline{K}_0\cdot\underline{r}} u_{\underline{k}_2+\underline{K}_0}(\underline{r}) = v_{\underline{k}_2}(\underline{r}). \tag{4.16}$$

This means $\phi_{\underline{k}_2}$ is a Bloch wave with wavevector $\underline{k}_2$.

• *We have proved that for any Bloch wavefunction with wavevector outside the first zone we can find an equivalent one which has wavevector inside the first zone.*

But we still have a notational difficulty. Each time we replace a wavefunction from outside the first zone with one inside it we have to give the new wavefunction a distinct name.[14] We can't call it $\psi_{\underline{k}}$ because there are already Bloch waves in the first zone. So when we replace one from outside with a new one inside we simply give the new wavefunction a new band index n. (Of course we could have called the new function $\phi_{\underline{k}}$, as we did above, but we soon run out of symbols.) For the free electron model we have the set of wavefunctions:

$$\psi_{\underline{k}},\ \underline{k} \in [\text{all k space}]. \tag{4.17}$$

[14] ψ and ϕ have a different k dependence and are therefore distinct functions.

In a crystal we have the set

$$\psi_{n\underline{k}},\ \underline{k} \in [1\,\mathrm{BZ}], n = 1, 2, 3 \cdots . \tag{4.18}$$

Band indices are usually allocated to ensure symmetry of the energy bands (i.e. such that $\epsilon_n(\underline{k}) = \epsilon_n(-\underline{k})$). We will see later that this helps us to visualize electron trajectories.

4.5 Solving the Schrödinger equation

We saw in the previous section that the wavefunctions for electrons moving in a periodic system are Bloch waves. That Bloch waves can be conveniently expanded as superpositions of plane waves leads one to a Fourier approach to solving the Schrödinger equation. We will see how this is done below, but remember that we are still *assuming* that we already know the self-consistent crystal potential U.

4.5.1 The Schrödinger equation in matrix form

Substituting the plane wave expansion of the Bloch wavefunction (Eq. 4.14 on the preceding page) into the one-electron Schrödinger equation (Eq. 4.3 on page 93) it follows that

$$\left\{ \frac{\hbar^2}{2m}(\underline{K} + \underline{k})^2 - \epsilon(\underline{k}) \right\} B^{\underline{k}}_{\underline{K}} + \sum_{\underline{K}'} U_{\underline{K}-\underline{K}'}\, B^{\underline{k}}_{\underline{K}'} = 0 \tag{4.19}$$

where the $B^{\underline{k}}$ are the plane wave coefficients, and the crystal potential $U(\underline{r})$ has been written as the Fourier series:

$$U(\underline{r}) = \sum_{\underline{K}} U_{\underline{K}}\, e^{i\underline{K}.\underline{r}}. \tag{4.20}$$

For each value of $\underline{k}$ [15] Eq. 4.19 gives an infinite set of equations (one for each reciprocal lattice vector $\underline{K}$). We can rewrite Eq. 4.19 in the form

$$\sum_{\underline{K}'} \left\{ \left[\frac{\hbar^2}{2m}(\underline{K}' + \underline{k})^2 - \epsilon(\underline{k}) \right] \delta_{\underline{K}\underline{K}'} + U_{\underline{K}-\underline{K}'} \right\} B^{\underline{k}}_{\underline{K}'} = 0 \tag{4.21}$$

which now looks a bit like a matrix equation. The term in the curly brackets corresponds to a matrix (with rows and columns numbered by $\underline{K}$ and $\underline{K}'$ respectively), while the $B^{\underline{k}}_{\underline{K}'}$ are the components of a vector. Non-trivial solutions of a matrix equation with the form $Mv = 0$ (where M is a matrix and v is a column vector) only occur when the determinant of the matrix $\|M\|$ is zero. In the present case we require

$$\left\| \left[\frac{\hbar^2}{2m}(\underline{K}' + \underline{k})^2 - \epsilon(\underline{k}) \right] \delta_{\underline{K}\underline{K}'} + U_{\underline{K}-\underline{K}'} \right\| = 0. \tag{4.22}$$

[15] In fact we need only consider an "irreducible wedge" of the first BZ. Remember from CMP that we always try to avoid wasting effort by considering two or more $\underline{k}$ points which are equivalent.

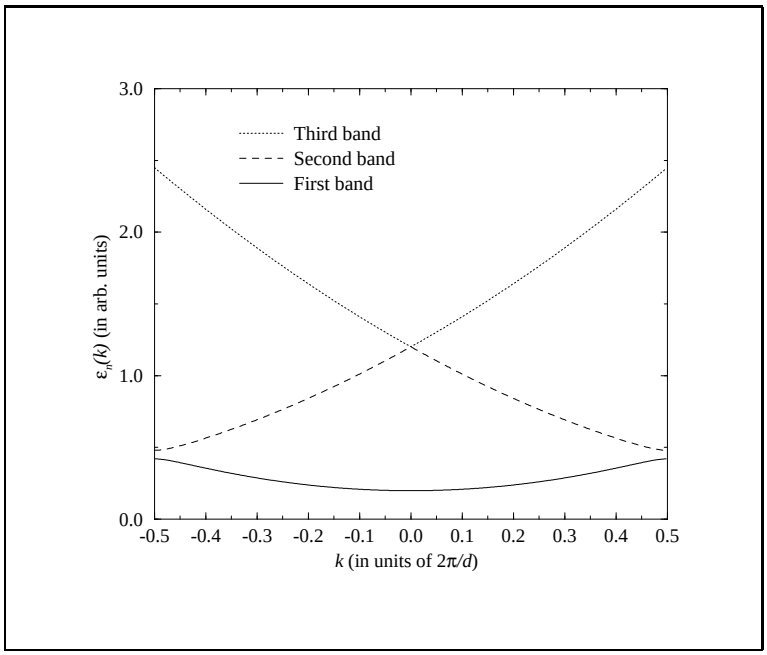

Figure 4.12: Solutions of the secular equation 4.24 on the following page.

We denote the energies which give this condition as $\epsilon_n(\underline{k})$, and the n subscript which distinguishes the different energies is the band index introduced above. For each band index we have a set of $B^{\underline{k}}$ coefficients which give us the wavefunction $\psi_{n\underline{k}}$. As we vary $\underline{k}$ in the first BZ the energies $\epsilon_n(\underline{k})$ vary smoothly, mapping out a series of curves - the "energy bands". Thus the problem of calculating energy bands is equivalent to solving the determinantal (sometimes called *secular*) equation 4.22.

4.5.2 The NFEM revisited

It should be obvious that the ease with which this can be done is determined by the size of the matrices, i.e. how many reciprocal lattice vectors are included in the plane wave expansion of the Bloch waves and the potential (Eq. 4.14 on page 107 and 4.20). If we keep only one (all the $B^{\underline{k}}_{\underline{K}}$ are zero except $B^{\underline{k}}_0$) then we just get the free electron result $\epsilon(\underline{k}) = (\hbar k)^2/2m$. A slightly more difficult case is the nearly free electron model (NFEM) in 1D which we considered earlier. If we keep the three reciprocal lattice vectors $K_{-1} = -2\pi/d, K_0 = 0, K_1 = 2\pi/d$, then Eq. 4.19 becomes the set

$$\begin{aligned}
\left[\frac{\hbar^2}{2m}(K_{-1}+k)^2 - \epsilon(k)\right] B^k_{-1} + U_0\, B^k_{-1} + U_{-1}\, B^k_0 &= 0 \\
\left[\frac{\hbar^2}{2m}(K_0+k)^2 - \epsilon(k)\right] B^k_0 + U_1\, B^k_{-1} + U_0\, B^k_0 + U_{-1}\, B^k_1 &= 0 \\
\left[\frac{\hbar^2}{2m}(K_1+k)^2 - \epsilon(k)\right] B^k_1 + U_1\, B^k_0 + U_0\, B^k_1 &= 0. \qquad (4.23)
\end{aligned}$$

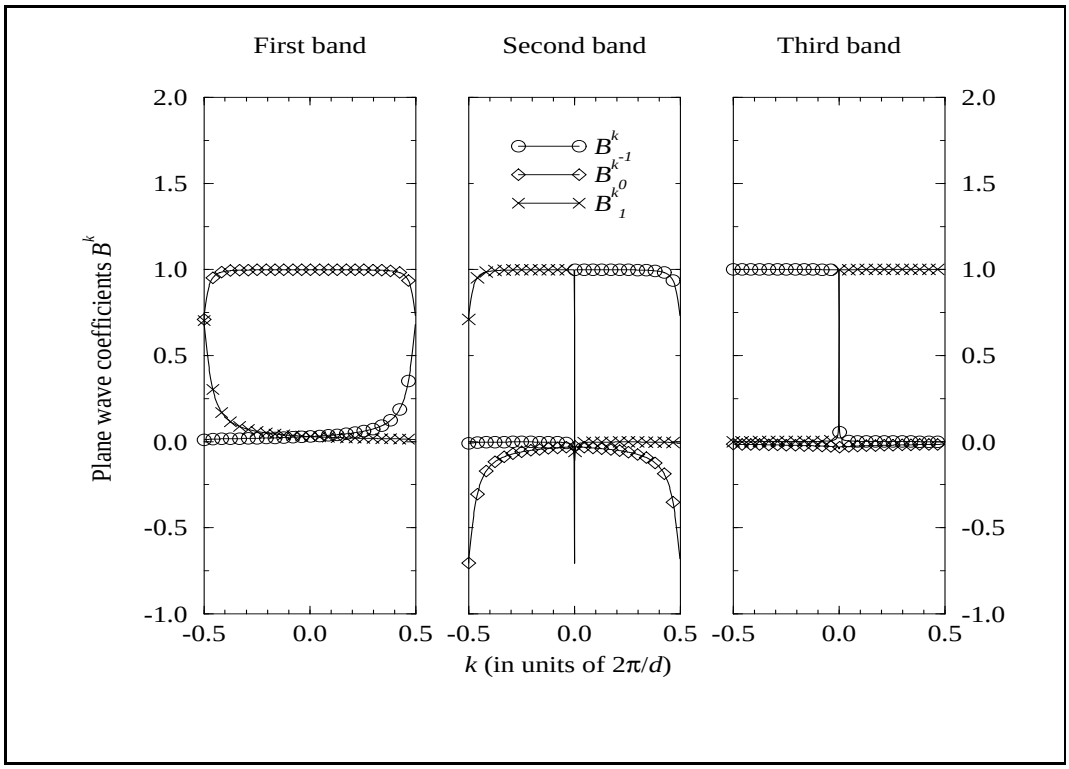

Figure 4.13: Plane wave coefficients for the bands shown in Fig. 4.12.

This can be arranged as the matrix equation

$$\begin{pmatrix} \frac{\hbar^2}{2m}(k+K_{-1})^2+U_0-\epsilon(k) & U_{-1} & 0 \\ U_1 & \frac{\hbar^2}{2m}(k+K_0)^2+U_0-\epsilon(k) & U_{-1} \\ 0 & U_1 & \frac{\hbar^2}{2m}(k+K_1)^2+U_0-\epsilon(k) \end{pmatrix} \begin{pmatrix} B^k_{-1} \\ B^k_0 \\ B^k_1 \end{pmatrix} = 0 \tag{4.24}$$

Setting the determinant to zero gives

$$\left[\frac{\hbar^2}{2m}(k+K_{-1})^2+U_0-\epsilon(k)\right] \times \left[\frac{\hbar^2}{2m}(k+K_0)^2+U_0-\epsilon(k)\right]$$

$$\times \left[\frac{\hbar^2}{2m}(k+K_1)^2+U_0-\epsilon(k)\right] -$$

$$U_1U_{-1}\left[\frac{\hbar^2}{2m}(k+K_{-1})^2+\frac{\hbar^2}{2m}(k+K_1)^2+2U_0-2\epsilon(k)\right]=0 \tag{4.25}$$

The energy bands satisfying this equation[16] are shown in Fig. 4.12 on page 109. As $\underline{k}$ varies we see three distinct curves, the energy bands, and we label each one with a different band index. For the first band we find that B^k_0 is the only significant plane wave component near the centre of the BZ i.e. we just have a

[16]The values $U(K_1) = U(K_{-1}) = -0.03, U(K_0) = 0.2$ (in the units shown in Fig. 4.12) have been used.

free electron with wavevector k. But near $\underline{k} \lesssim \pi/d$ we find $B_0^k \approx B_{-1}^k \gg B_1^k$ for the first band and $-B_0^k \approx B_{-1}^k \gg B_1^k$ for the second band. This is precisely the NFEM we considered earlier, except here we did things properly. Note that we only got the first three bands because a 3×3 matrix can only have three independent eigenvectors. Notice also that no band gap emerged between the second and third bands, simply because the corresponding Fourier coeffiecient of the crystal potential was implicitly zero. Including the zeroth-order Fourier component of the potential (i.e. a constant) just shifted the zero of the energy bands.

4.6 Real materials

We have seen how a weak crystal potential affects free electrons, but one must ask how relevant the NFEM is to real materials. To rephrase this question we could ask "What is U really like in real solids?" There is a very good indication in the ionization energies of the elements. (The ionization energy is the energy required to pull of an electron off an atom.) Although Li, Na, K atoms are similar to the H atom in having a single outer electron, their ionization energies are considerably less than 13.6 eV. Furthermore we know that the trend is for the ionization energy to increase down the group. So when we notice the conduction electrons in metallic sodium appear to pay little regard to the charged ions, we are in a sense merely rediscovering the same physics beneath the ionization energy trend. This same trend is observed in Group 4 of the periodic table. The crystal potential is strong at the top of the Group so here we find electrical insulators. The ionization energy and the band gap each steadily decrease down the group until we get to tin and lead which are metals.

4.6.1 The problem with plane waves

The spirit of the nearly free electron model can be summed up as "try to get away with as few plane waves as possible". It turns out that the alkali metals are quite NFE-like. This is in itself quite hard to explain, but our aim here is quite ambitious - we want a scheme capable of describing *any* (periodic) solid, and with quantitative accuracy. We have no reason to suspect that the crystal potential is always weak. We have seen already the general form of electron wavefunctions in a crystal (i.e. the Bloch form) but let's try to see what these actaully look like for real solids.

The "core" electrons (the ones deep within atoms which do not particupate in bonding) in a crystal must obey Bloch's theorem. As seen below this can be achieved by making linear combination of truly atomic core wavefunctions, which we will assume are known. It follows that while their wavefunctions are concentrated near the nuclei, they are modulated by a plane wave factor, as shown in Fig. 4.14. A typical valence wavefunction is also shown in Fig. 4.14. Since the eigenstates of a Hamiltonian must be orthogonal to each other, the valence wavefunction displays rapid oscillations near the nucleus to ensure orthogonality with all the core electrons.[17] Although the valence wavefunction

[17]The wiggles can also be qualitatively understood on energetic grounds: Both core and valence electrons experience an extremely large and negative potential energy when near the nucleus. Since core levels have lower energy than valence levels this means that while near

is quite plane wave like between each atomic sites, a Fourier expansion that does justice to the wiggles would require far too many plane waves for practical calculations to be attempted.

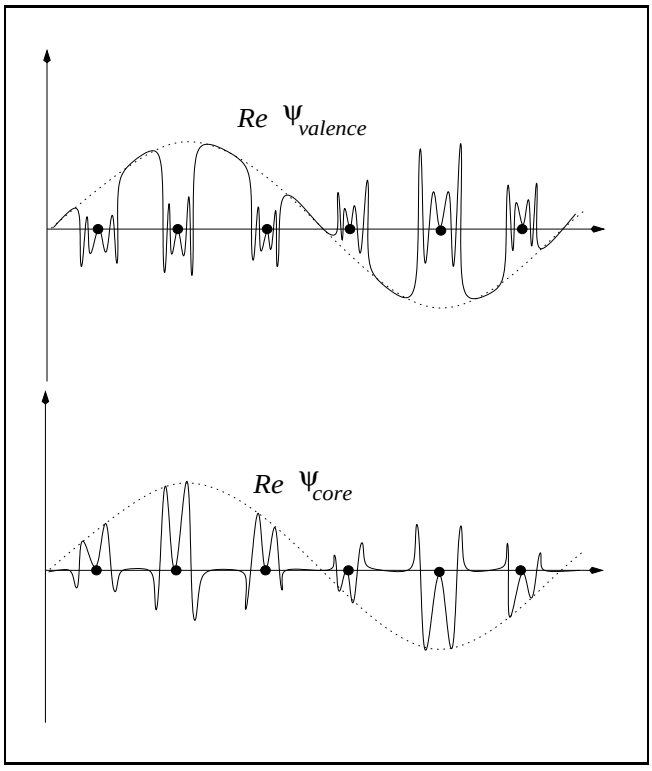

Figure 4.14: Characteristic spatial dependence of valence and core level wavefunctions in a solid.

Since the core electrons are not involved in the formation of chemical bonds they are not of primary interest to us. Rather, we are concentrating our efforts on finding the valence Bloch functions. Thus we have regarded the crystal potential U which enters the one-electron Schrödinger equation as a combination of a repulsive term between pairs of valence electrons and a net attractive term due to the interaction between each valence electron and each atomic core (i.e. nucleus plus core electrons). The first term, the electron-electron repulsion, is a many body term which we have not yet addressed. We will come to that shortly. At the moment we must concentrate on the problem arising from the electron-core term: how to accurately expand the wavefunctions yet retain a manageable number of basis functions.[18]

The problem can be alleviated by modifying either

the nucleus the valence levels must have the larger kinetic energy. Since the velocity operator is $-(i\hbar/m)\,\nabla$ it follows that valence levels have the more rapidly varying wavefunction when near the nucleus.

[18] We are assuming that we already know the atomic core wavefunctions. If these are largely unperturbed in the crystal, save for the phase factors required to satisfy Bloch's theorem, then we know in advance the core Bloch functions. This means that the (valence) electron-ion potential term is not a many body potential.

(i) the basis functions, or
(ii) the potential used in the Schrödinger equation.
We now consider some of the dominant methods currently in use.

4.6.2 The orthogonalized plane wave method

In the previous subsection we noted the fundamentally different character of what we call *core* and *valence* electrons. The former are those electrons in inner shells which are tightly bound to the nucleus and which are therefore largely oblivious to the local chemical environment (i.e. they are insensitive to the chemical bonds formed by the atom on which they are trapped). Valence wavefunctions are substantially redistributed when atoms bond together, but they must still remain orthogonal to the core states, a condition which requires them to be wiggly in the core region.

While these considerations make plane waves unsuitable as a basis set for expanding Bloch waves, Herring (1940) devised a simple modification of this idea. Herring's method, the *orthogonalized plane wave method* (OPW), is based on the suggestion that a wave function of the form

$$\phi_{\underline{k}}(\underline{r}) = e^{i\underline{k}.\underline{r}} + \sum_c A_c^{\underline{k}} \psi_{\underline{k}}^c(\underline{r}) \tag{4.26}$$

where $\psi_{\underline{k}}^c$ are the core level wavefunctions of the solid and the sum is over all core levels, is likely to give a rather good representation of a valence wavefunction *in all regions of space*. The OPWs $\phi_{\underline{k}}$ are required to be orthogonal to the core levels, a condition implying

$$A_c^{\underline{k}} = -\int d\underline{r} \left[\psi_{\underline{k}}^c(\underline{r})\right]^* e^{i\underline{k}.\underline{r}}. \tag{4.27}$$

Although the core levels (by definition) are not involved in the bonding of solids, we cannot just use atomic core level wavefunctions in these expressions since they don't satisfy Bloch's theorem. But we can easily give them Bloch form by using the superposition:

$$\psi_{\underline{k}}^c(\underline{r}) = \sum_{\underline{R}} e^{i\underline{k}.\underline{R}} \psi^c(\underline{r} - \underline{R}) \tag{4.28}$$

where $\psi^c(\underline{r})$ is the corresponding level in a free atom.[19] Since the core wavefunctions are localized around atomic sites, as shown in Fig. 4.14, each OPW has both a component which oscillates near the nuclei as well as a plane wave component where the potential is much weaker. On this basis one expects that an expansion of the valence Bloch waves of the form

$$\psi_{\underline{k}}(\underline{r}) = \sum_{\underline{K}} C_{\underline{K}}^{\underline{k}} \phi_{\underline{k}+\underline{K}} \tag{4.29}$$

will yield a good approximation even with only a few components.

[19] In a problem sheet we will soon show that strictly speaking the core states of solids form flat energy bands which are fully occupied.

For comparison with 4.19 on page 108, where the potential and wavefunctions were written as plane wave expansions, the Schrödinger equation for the OPW method can be written

$$\left[\frac{\hbar^2}{2m}(\underline{K}+\underline{k})^2 - \epsilon(\underline{k})\right] C^{\underline{k}}_{\underline{K}} + \sum_{\underline{K}'} W_{\underline{K},\underline{K}'}\, C^{\underline{k}}_{\underline{K}'} = 0. \tag{4.30}$$

It is tempting to imagine that this expression describes plane wave coefficients for an electron moving in the potential W. In fact the OPW matrix elements $W_{\underline{K},\underline{K}'}$ are much smaller than the equivalent terms $U_{\underline{K}-\underline{K}'}$ in Eq. 4.19. The physical interpretation of this is as follows: the orthogonalization terms introduced in the OPW construction tend to cancel the potential well of the nuclei.[20] In other words the Pauli repulsion between the core and valence electrons prevents the valence electrons "seeing" the full nuclear potential. Indeed for the free electron metals the valence electrons hardly even notice that there are nuclei around.

Thus the OPW method clarifies why the alkali metals are free electron like. It also prompts the question: is there a potential (a *pseudopotential*) that gives the same energy bands as the true crystal potential but which is "softer" and so easier to represent as a plane wave expansion?

4.6.3 Pseudopotentials

The way the pseudopotential idea is usually implemented is as follows. We take a free atom, for which the electronic structure is well known, and calculate its valence level wavefunctions and energies. We wish to cut out the nucleus and core electrons and replace them with a pseudopotential which gives the same valence level energies and the same valence wavefunctions (beyond the core radius). As hinted at above, one can reformulate the OPW method to provide a pseudopotential recipe, but there are better ways. Having determined a pseudopotential that works for an atom we can use it for solids. The smoothness of both pseudopotentials and pseudowavefunctions, such as those shown in Fig. 4.15 for Si,[21] means that they can be efficiently expanded using plane waves.

Notice that the pseudopotential method (by construction) only gives a description of the electronic states involved in the chemical bonding within a solid. The statement that core levels do not participate in chemical bonding tends to suggest that they are completely irrelevant, but it is vital to appreciate that this is not the case. We can see this by comparing the solids carbon, silicon, germanium, tin and lead. Each of these elements is in group 4 of the Periodic Table and each has four valence electrons, but their properties (electrical, chemical, mechanical, optical) are strikingly different. How can this be? The answer is their cores are very different. The core electrons themselves are of little interest but they play a decisive role in determining the behaviour of the valence electrons. We will return to this important point later on.

[20] We will demonstrate this explicitly in a Problem Sheet.

[21] Note that the pseudopotential in Fig. 4.15 is weaker than the potential for the 4+ ion so it's not simply a matter of cancelling some protons in the nucleus with some core electrons.

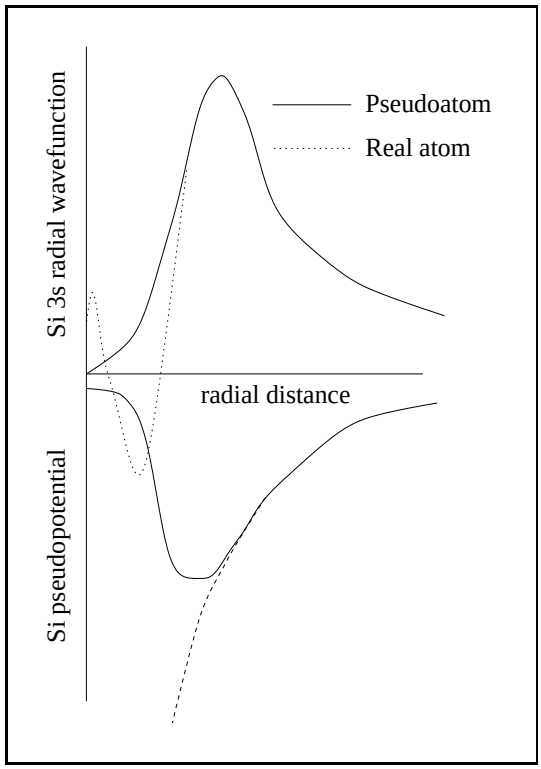

Figure 4.15: A pseudopotential and pseudowavefunction for Si. Also shown is the Coulomb potential of a 4+ ion (dashed line).

4.6.4 Augmented plane waves

One can simplify the potential in a different way. $U(\underline{r})$ is assumed to be spherically symmetrical within some radius (called the *muffin tin* (MT) radius), and constant in the interstitial region, as shown in Fig. 4.16. It is then natural to construct an *Augmented plane wave* as a spherically symmetric (atomic-like) orbital within the MT sphere matched to a plane wave in the interstitial region. APWs offer another efficient basis set for the expansion of valence wavefunctions in solids, and incidentally, the APW method can also be cast into pseudopotential form.

4.6.5 Multi-atom unit cells

In a footnote at the start of §1 we decided to consider band calculations for a single atom unit cell. Here we briefly mention how things go with a multi-atom unit cell. The total potential U is written as a sum over the atoms in the unit cell. The Fourier components of this potential then take precisely the same form as the structure factors we saw in x-ray diffraction - each is a sum over the j atoms in a unit cell of an "atomic form factor" (determined by the type of atom j), multiplied by a phase factor (determined by the position of atom j).

Bloch's theorem applies only to periodic solids, but can we do anything for non-periodic solids? We can think of non-periodic solids as crystals with an

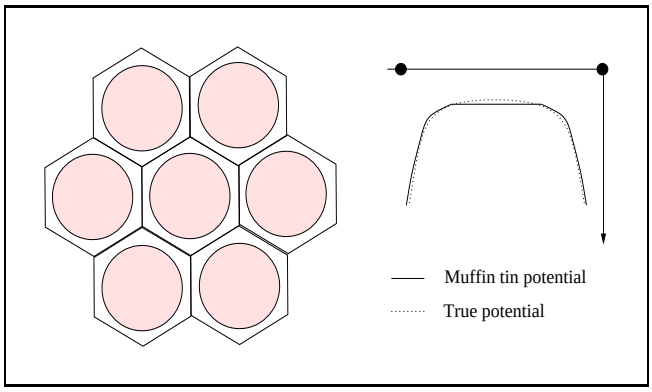

Figure 4.16: The shaded areas denote the spherically symmetrically region within MT spheres. The MT potential, shown on the right, is constant in the interstitial regions.

infinite unit cell. This notion is, of itself, little use but it is useful to go a step further by saying that the unit cell is merely "large". We can then apply Bloch's theorem to this large *supercell*. These days *ab initio* calculations can be performed for cells containing over 1000 atoms, so it is likely that for many purposes the supercell approximation is reasonable. This is just a trick, however. We will try something with more physical insight later, if we have time.

We have now seen the general form of the Schrödinger equation and its eigenstates in a periodic solid and we have seen methods designed to compute them. To implement these methods we need to be specific about the potential U. We have seen that the interaction between the ion cores and the valence electrons is amenable to simplification by the pseudopotential method, for example. The missing ingredient is the interactions *between* the valence electrons. This is the subject of the following section.

4.7 The electron-electron interaction

In §4.1 we noted that the Schrödinger equation for the electronic wavefunction of a solid represents an insoluble many body problem. The one electron approximation (OEA) simplifies the situation considerably, yielding a manageable set of equations each describing only a single electron. The development of a reliable OEA has been one of the most important achievements in solid state physics (and chemistry) in the last few decades. In this section we have to be more specific about the nature of this approximation, how it might be implemented, and how it can be justified.

For notational convenience we will make two unimportant simplifications in this section. Firstly, we again assume a single atom basis. Secondly, we adopt the pseudopotential method, assuming the interaction between a valence electron and a particular ion core can be represented by some function U^{PS}. We are lumping the core electrons and nuclei together. We are left with the task of accounting for the interactions between valence electrons.

4.7.1 The Hartree potential

The fundamental idea of the OEA is to associate a specific wavefunction ψ with each electron. These ψ, each a function of the coordinates of a single electron, are then used to form Ψ, the wavefunction for the system as a whole. If electrons did not interact with each other then the position of any particular electron would not affect the position of any other.[22] This means that the motion of the electrons is statistically independent and we can write

$$\Psi(\underline{r}_1, \underline{r}_2, \cdots, \underline{r}_N) = \psi_1(\underline{r}_1)\psi_2(\underline{r}_2)\cdots\psi_N(\underline{r}_N) = \prod_1^N \psi_i(\underline{r}_i) \tag{4.31}$$

where the ψ_i are the N single electron wavefunctions with the lowest energy, filled up in accord with the Pauli principle.[23]

Aware as we are of the Coulombic force between electrical charges, we must regard the notion of non-interacting electrons as totally inadequate. If we are to insist upon retaining a wavefunction of the form given by Eq. 4.31 then we can at best account for electron-electron repulsion in an average way. We can do this by first building up the (valence) charge density of the system

$$n(\underline{r}) = \sum_i |\psi_i(r)|^2. \tag{4.32}$$

This quantity tells us where the electrons can be found on average, but tells us nothing about where any particular electron is located at any particular instant. We now assume that a particular (valence) electron, "electron one" say, in the system feels the repulsion that would arise if all the other electrons adopted the average distribution $n(\underline{r})$. This average electron-electron repulsion is usually

[22] You may object to this on symmetry grounds in relation to the Pauli principle, but we'll come to that soon.

[23] Remember that we are dealing here with the valence electrons only.

referred to as the *Hartree potential* U^H and is given by[24]

$$U^H(\underline{r}) = e^2 \int \frac{n(\underline{r}')}{4\pi\varepsilon_0 \, |\underline{r} - \underline{r}'|} d\underline{r}'. \tag{4.33}$$

The one electron Schrödinger equations are now of the form

$$-\frac{\hbar^2}{2m}\nabla^2\psi_i(\underline{r}) + \left\{\sum_R U^{ps}(\underline{r} - \underline{R})\right\}\psi_i(\underline{r}) + \left\{\sum_j \int d\underline{r}' \frac{e^2|\psi_j(\underline{r}')|^2}{4\pi\varepsilon_0 \, |\underline{r} - \underline{r}'|}\right\}\psi_i(\underline{r}) =$$

$$-\frac{\hbar^2}{2m}\nabla^2\psi_i(\underline{r}) + \left\{\sum_R U^{ps}(\underline{r} - \underline{R})\right\}\psi_i(\underline{r}) + U^H(\underline{r})\psi_i(\underline{r}) = \epsilon_i\psi_i(\underline{r}) \tag{4.34}$$

known as the *Hartree equations*. There are some appealing features to this expression.

It "looks right". In fact we have written down the Hartree equations largely on the basis of plausibility.[25]

On the down side, the Hartree equations are coupled - the equation for electron i depends on the wavefunction of all the other electrons. The equations must be solved *iteratively* starting with a suitable guess for the wavefunctions and hence Hartree potential. We then solve the Hartree equations to obtain new wavefunctions and a new Hartree potential. We repeat this cycle until the wavefunctions and potential are *self-consistent*. For this reason the Hartree method is called the *self-consistent field* method.

The Hartree method is a good starting point for the discussion of electron-electron interactions, but we must always bear in mind its deep-rooted shortcoming: we have assumed that at any particular instant an electron does not care where any of the others actually are - only where they are *on average*. The technical term for this is the *neglect of correlation*. In reality electron motions *are* correlated for two reasons:

1. Coulomb Correlation
Since electrons repel each other they will keep as far apart from each other as possible. If we take the example of the hydrogen molecule from we can easily accept that at any instant it would be highly unlikely for both electrons to be "on" the same atom. If we know where electron one is then we can predict with good certainty where electron two is, just on the basis of electrostatics. In the Hartree approximation we assume that any particular electron does not know where any other electron is at any moment, but only their *time-averaged* positions. As a result the Hartree approximation allows electrons to occasionally come very close to each other, a configuration with a high energy cost. Thus the Hartree approximation slightly overestimates electron-electron repulsions, but it is rather hard to systematically improve the method.

[24] We are being a bit sloppy here since we have accidentally allowed each electron to repel itself. For solids, including the repulsion due to one extra delocalized charge among say 10^{23} others makes a negligible difference, but this is not true for an atom.

[25] In a Quantum Mechanics course you have derived this result from the variational principle. The Hartree equations produce the best possible solution to the Schrödinger equation that can be written in the form of Eq. 4.31.

2. Exchange
In the discussion above we have been speaking about "electron one" and so on as if we could distinguish them. Of course we can't and this gives us a second source of correlation usually referred to as the *exchange* effect. The Hartree wavefunction given by a product of one electron wavefunctions (Eq. 4.31) is consistent with the usual statement of the Pauli principle: each $\psi(\underline{r})$ cannot be multiply occupied, apart from a double occupation due to spin degeneracy, but there is a stronger requirement. Since electrons are not distinguishable and have half-integer spin, the wavefunction of an N electron system must change sign on interchange of any two of its particles:

$$\Psi(\underline{r}_1 s_1, \cdots, \underline{r}_i s_i, \cdots, \underline{r}_j s_j, \cdots, \underline{r}_N s_N) =$$
$$-\Psi(\underline{r}_1 s_1, \cdots, \underline{r}_j s_j, \cdots, \underline{r}_i s_i, \cdots, \underline{r}_N s_N) \tag{4.35}$$

where s is spin. Writing each one electron wavefunction as the product of a "space function" and a "spin function", it can be shown that this fundamental requirement introduces a special form of electron correlation: electrons with parallel spins tend to avoid each other. Each electron is said to carry around an *exchange hole*, a region in which other electrons with the same spin are excluded.

4.7.2 Hartree-Fock

We can account for antisymmetry by writing Ψ as a *Slater determinant* of one-electron wavefunctions. This construction leads to the *Hartree-Fock* one electron equations:

$$-\frac{\hbar^2}{2m}\nabla^2\psi_i(\underline{r}) + \left\{\sum_R U^{ps}(\underline{r}-\underline{R})\right\}\psi_i(\underline{r}) + \sum_j \int d\underline{r}' \frac{e^2}{4\pi\varepsilon_0\,|\underline{r}-\underline{r}'|}\psi_j^*(\underline{r}')\psi_j(\underline{r}')\psi_i(\underline{r})$$
$$-\sum_j \int d\underline{r}' \frac{e^2}{4\pi\varepsilon_0\,|\underline{r}-\underline{r}'|}\psi_j^*(\underline{r}')\psi_i(\underline{r}')\psi_j(\underline{r})\delta_{s_i s_j}$$
$$-\frac{\hbar^2}{2m}\nabla^2\psi_i(\underline{r}) + \left\{\sum_R U^{ps}(\underline{r}-\underline{R})\right\}\psi(\underline{r}) + \int U^{HF}(\underline{r},\underline{r}')\psi_i(\underline{r}') = \epsilon_i\psi_i(\underline{r}). \tag{4.36}$$

This expression is just Eq. 4.34 with an exchange term tagged onto the end of the right hand side. It is now explicit that the antisymmetry of Ψ produces a spatial separation of electrons with the same spin, effectively reducing the net electron-electron repulsion.

Like the Hartree case, solution of the Hartree-Fock (HF) equations requires iteration to self-consistency but the equations are somewhat more difficult to solve since the HF operator is *non-local*. While the Hartree equation (Eq. 4.34) for $\psi_i(\underline{r})$ can be solved for each $\underline{r}$ separately, the corresponding HF equation contains the value of ψ_i *at all other positions* $\underline{r}'$. This is a non-trivial complication. At this point we must resort to a system simple enough to yield soluble HF equations, and it is to the free electron gas to which we must turn. Though apparently a retrograde step, consideration of the free electron gas played a pivotal role in the historical development of the theory of many-electron systems, and it is quite surprising that it remains very much at the heart of today's most successful electronic structure calculations.

4.7.3 HF theory of the free electron gas

The first question we must address is: where are the ion cores? The free electron model assumes the electrons to experience a constant potential, which is often simply set to zero. For this to be the case the positive charge of the ion cores must be uniformly distributed throughout space, and this is usually referred to as the *jellium* model. Now can we just "see" the solutions to the HF equations? We might guess that plane waves will do, since these ensure a uniform electron density and hence a uniform potential. Substituting plane waves into the HF equations leads to eigenenergies satisfying

$$\epsilon_{\underline{k}} = \epsilon^0_{\underline{k}} - \frac{2k_f e^2}{\pi}\left\{\frac{1}{2} + \frac{k_f^2 - k^2}{4kk_f}\ln|\frac{k_f + k}{k_f - k}|\right\} \tag{4.37}$$

where $\epsilon^0_{\underline{k}} = \hbar^2k^2/(2m)$, and k_f is the Fermi wavevector. Clearly the first term on the right hand side is the electronic kinetic energy term familiar to us from CMP. Since an elemental charge in jellium will experience equal and opposite electrostatic forces from the positive ion cores and negative electrons, both of which are homogeneous distributions, the complicated looking second term must come the exchange term in the HF equation. Although the electron density is homogeneous on average, at any particular moment an electron will be surrounded by an exchange hole. The effect of this correlation effect is to lower the electron's energy since there is now more positive charge than negative in the vicinity of the electron. When electron-electron interactions are explicitly considered, the free electron gas is often referred to as the *homogeneous electron gas*.

We are now starting to address a long-running embarrassment. Earlier we claimed that many metals are free-electron-like, while noting that the free electron model contains only the kinetic energy term ϵ^0_k, which is positive. It is not possible to understand the cohesion of metals within this framework. What holds a metal together? A free electron metal should spontaneously expand so that its energy can be lowered, the energy tending to zero as the electron density approaches zero. In fact better still, the metal should spontaneously form individual atoms, each of which would have *negative* energy. The HF result for the homogeneous electron gas shows that the exchange energy stabilizes a metal.

We can use the HF energy bands from Eq. 4.37 to compute the total energy of the homogeneous electron gas. This is not simply a matter of summing $\epsilon_{\underline{k}}$ since the exchange term is a "two-body" interaction energy. If we were to simply sum $\epsilon_{\underline{k}}$ for $k \leq k_f$ then we would get the kinetic energy part correct but we would double count all the electron-electron interaction terms. The total energy is therefore

$$E = 2\left\{\sum_{k\leq k_f}\epsilon^0_{\underline{k}} - \frac{1}{2}\times\frac{2k_f e^2}{\pi}\sum_{k\leq k_f}\left[\frac{1}{2} + \frac{k_f^2 - k^2}{4kk_f}\ln|\frac{k_f + k}{k_f - k}|\right]\right\} = N\left[\frac{3}{5}\epsilon_f - \frac{3}{4}\frac{e^2k_f}{\pi}\right]. \tag{4.38}$$

It is conventional to express this result in terms of r_s, the radius of the sphere whose volume is equal to the volume per conduction electron in a metal:

$$r_s = \left(\frac{3}{4\pi n}\right)^{1/3} \tag{4.39}$$

where n is the number of conduction electrons per unit volume. Making this substitution we obtain

$$\frac{E}{N} = \frac{30.1}{(r_s/a_0)^2} - \frac{12.5}{r_s/a_0} \tag{4.40}$$

where $a_0 = 0.53$ Å is the Bohr radius. Using sensible values of r_s/a_0 (e.g. 3) we obtain an energy of the order of ~ -1 eV per electron. This number seems reasonable but it is now clear that any quantitative study of the cohesion of metals must include a good treatment of the rather awkward many-body effects. We should also note that the above expression does not distinguish one metal from another.

There is one more useful observation we can make about the HF treatment of the homogeneous electron gas. Given the extreme difficulty of solving the HF equations for real systems, Slater struck upon the idea of adding a local correction to the Hartree equations which would give an approximate treatment of the exchange effect. Noticing that the exchange contribution to the total energy of the homogeneous electron gas scales with the inverse of r_s, and hence as $n^{1/3}$, Slater suggested that for real materials one might add a local potential $U^X \propto n(\underline{r})^{-1/3}$ to the Hartree potential, and this became known as the Hartree-Fock-Slater method. There followed some dispute regarding the appropriate weighting of the approximate exchange term. It transpired that the "correct" weight did not give optimum results as compared with experimental data. Slater responded by writing the exchange term as

$$U^X(\underline{r}) = X_\alpha n(\underline{r})^{1/3} \tag{4.41}$$

where X_α is a free parameter to be varied at one's discretion, and this approach became known as the X_α method.

This sounds quite unsatisfactory, but the idea of focussing on the electron density (and in particular on the *local* electron density) was to be highly significant.

4.7.4 Density functional theory

The Hohenberg-Kohn theorem

The ground state energy of a system is a unique functional of the electron density.

This is the Hohenberg-Kohn theorem(1964). Kohn and Sham (1965) went on to show that it is then possible to map exactly the problem of the interacting electron gas onto that of a set of self-consistent single particles moving in an effective potential:

$$\left\{ -\frac{\hbar^2}{2m}\nabla^2 + \sum_{\underline{R}} U^{ps}(\underline{r} - \underline{R}) + U^H(\underline{r}) + U^{xc}(\underline{r}) \right\} \psi_i(\underline{r}) = \epsilon_i \psi_i(\underline{r}) \tag{4.42}$$

where $U^{xc}(\underline{r}) = \delta E^{xc}[n(\underline{r})]/\delta n(\underline{r})$, and E^{xc} is the "exchange-correlation functional". These are the *Kohn-Sham* equations and the approach generally is referred to as *density functional theory*. It may seem as though we have just

bundled together all the difficult terms together into this mysterious exchange-correlation term. To some extent that is true, but it is a huge step to be able to show that we are *allowed* to reduce the many body problem to the Kohn-Sham one-electron equations.

The local density approximation

If the exchange-correlation energy functional were known we would have a scheme for computing exactly the energy and charge density of a solid. Unfortunately, the Hohenberg-Kohn theorem tells us such a functional exists and that it is unique, but it doesn't tell us what the functional actually is! The most widely used approach is to use the *local density approximation* (LDA). The LDA assumes that the exchange correlation potential at point $\underline{r}$ in a solid is given by that of a homogeneous electron gas which has the same density as the real system at the point concerned.[26] It is an approximation, but the LDA makes DFT calculations practical, and it seems to work pretty well.[27] In fact the vast majority of electronic structure calculations these days employ DFT in the LDA.[28] There are no theorems or *a priori* arguments to explain why DFT in the LDA should work. Nonetheless it does, and a widespread acceptance of the method has emerged on this basis over the last decade or so. Although we have skipped over some of the details, we have now seen how band structure calculations are done today. The steps are summarized in Fig. 4.17.

Incidentally, we can now see why an "incorrect" value of X_α gives optimum results. The "error" is accidentally accounting for correlation as well as exchange. The X_α method is therefore just a very crude version of density functional theory.

Limitations

We should briefly remind ourselves of the limitations of the application of density functional theory. Without being exhaustive, we can point to
1. The Hohenberg-Kohn theorem relates only the energy of the *ground state* of a system of electrons to its charge density. Quite often we will be concerned with excitations of the ground state. Formally, density functional theory is not valid in this case, but often it seems to still work.
2. We can never perform exact density functional calculations. We must always make some guess at the real exchange-correlation potential.
3. This is a subtle point. The Kohn-Sham equations are mathematical constructs which describe fictitious particles, and not necessarily real electrons. The only requirement is that the fictitious particles give the correct ground state energy and charge density. We have no right, therefore, to expect the spectrum of one-electron levels in a density functional calculation to resemble reality in any way. Despite points 1 and 3, density functional calculations often give spectra of one-electron levels which resemble experimental excitation spectra quite closely.

Let's now look at just a few examples of real materials.

[26] Exact results are available for an electron gas.

[27] In fact it worked much better than its proponents expected: Kohn and Sham stated "we do not expect an accurate description of chemical bonding".

[28] This accounts in part for the recent award of a Nobel prize to Kohn.

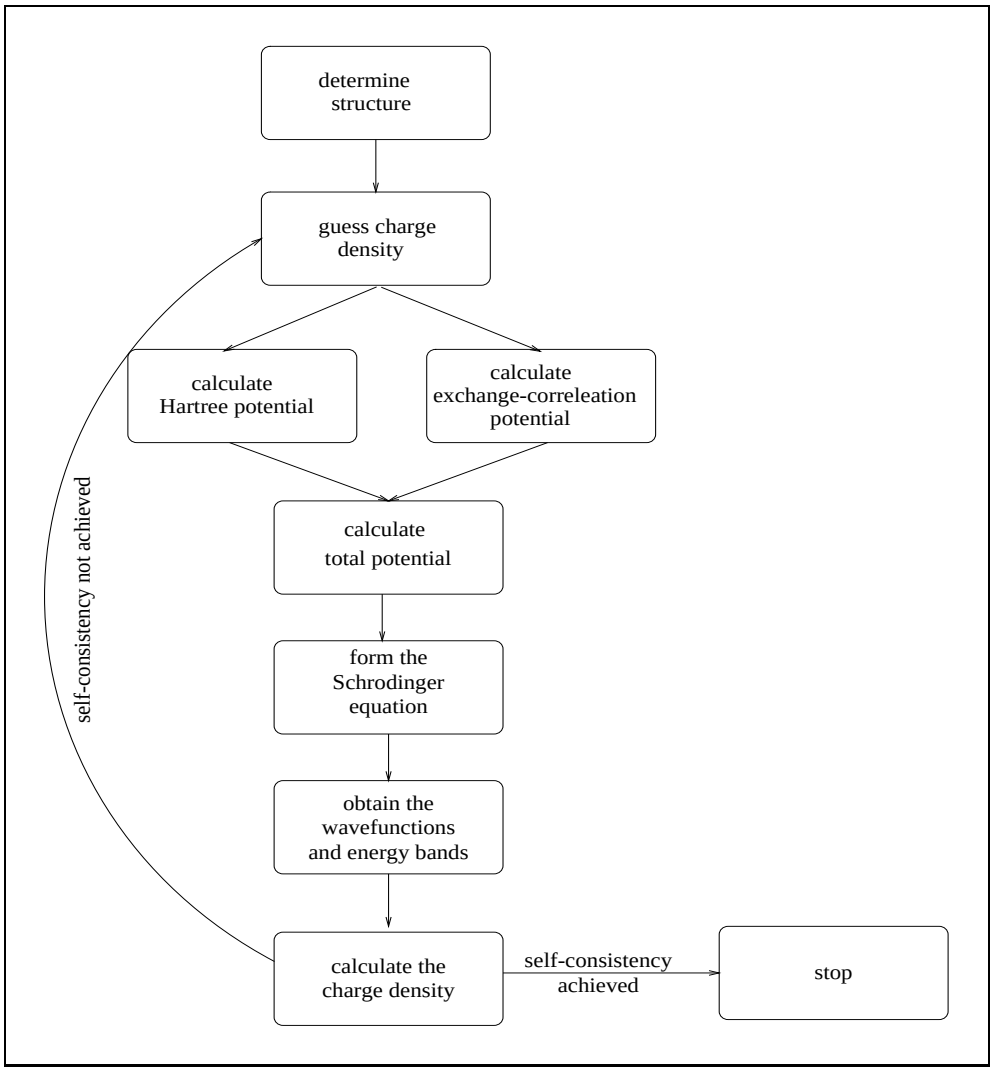

Figure 4.17: Flow diagram illustrating the steps of an *ab initio* electron structure calculation.

The alkali metals

These metals all adopt the bcc crystal structure. We can view the bcc crystal structure as a bcc Bravais lattice with a one atom basis and with primitive vectors $\underline{a} = L/2\,(-\underline{\hat{x}} + \underline{\hat{y}} + \underline{\hat{z}})$, $\underline{b} = L/2\,(\underline{\hat{x}} - \underline{\hat{y}} + \underline{\hat{z}})$, and $\underline{c} = L/2\,(\underline{\hat{x}} + \underline{\hat{y}} - \underline{\hat{z}})$, where L is the lattice constant (defined to be the length of the sides of the conventional bcc unit cell), and $\underline{\hat{x}}, \underline{\hat{y}}, \underline{\hat{z}}$ are unit cartesian vectors. The reciprocal of this lattice is an fcc lattice with lattice parameter $4\pi/L$.

In a truely free electron metal the surfaces of constant energy are concentric spheres. Experimental determination of the Fermi surfaces of the alkali metals reveal deviations from sphericity of only $\sim 0.1\%$. We can attribute this outcome to the fact that the crystal potentials in these metals are weak. Nonetheless one might have expected band gaps at BZ boundaries as seen in the NFEM. This is indeed the case, but since each alkali atom contributes only 1 electron per atom to the electron gas (and so 1 electron per primitive bcc unit cell) there are only sufficient electrons to half fill the first BZ, and so only the free electron like portions of the energy bands are occupied.

We saw in Problem sheet 1 that the conduction electron density in a free electron gas n is related to the Fermi energy ϵ_f by

$$n = \frac{2}{3}\frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2} {\epsilon_f}^{3/2} = \frac{1}{3\pi^2}{k_f}^3 \tag{4.43}$$

where k_f is the Fermi wavevector, and so the radius of the Fermi sphere can be expressed

$$k_f = \frac{(6\pi^2)^{1/3}}{L} = \left(\frac{3}{4\pi}\right)^{1/3}\frac{2\pi}{L} = 0.62\frac{2\pi}{L}. \tag{4.44}$$

The shortest distance from the centre of the first BZ to a zone face, which we will call k_{BZ}, is (see Fig. 4.18)

$$k_{BZ} = \Gamma\mathrm{N} = \frac{2\pi}{L}\sqrt{\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2} = 0.707\frac{2\pi}{L}. \tag{4.45}$$

From Eq. 4.44 and 4.45 we see that the Fermi sphere of an alkali metal is entirely contained within the first BZ. Evidently the crystal potential does not distort the sphere appreciably.

Group 4

Diamond is a very good insulator, Si and Ge are semiconductors, Sn is a semiconductor or a metal, and Pb is metallic. As we go down group 4, the band gap decreases from 5.4 eV in diamond, 1.1 eV in Si, 0.7 eV in Ge, and disappears altogether in Sn. There is also a structural trend - the insulators and semiconductors adopt the diamond crystal structure in which each atom forms bonds with four nearest neighbours, while the metals prefer more close packing.

A minor point which we might also mention here is that although the band gap of Si is 1.1 eV, its threshold for optical absorption is 2.5 eV.

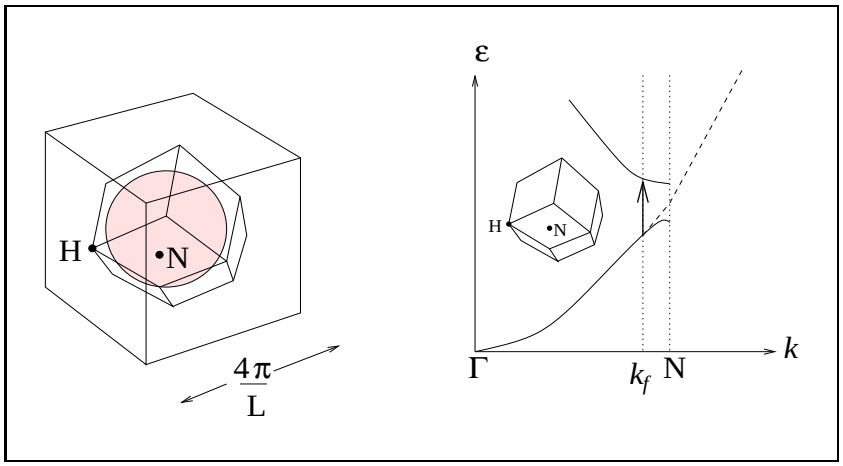

Figure 4.18: The free electron Fermi sphere of a monovalent bcc metal shown within the primitive unit cell of the reciprocal lattice. The H points lie at the centre of the faces of the conventional cube, the N points are on the midpoints of lines joining two H points, and the Γ point is at the centre of the cube. H and N lie on the first BZ boundary. Also shown is the energy bands for a monovalent bcc metal in the nearly free electron approximation. Although a band gap has opened up on the BZ boundary, the *occupied* bands are not significantly modified from free electron form. (We will discuss the arrow later.)

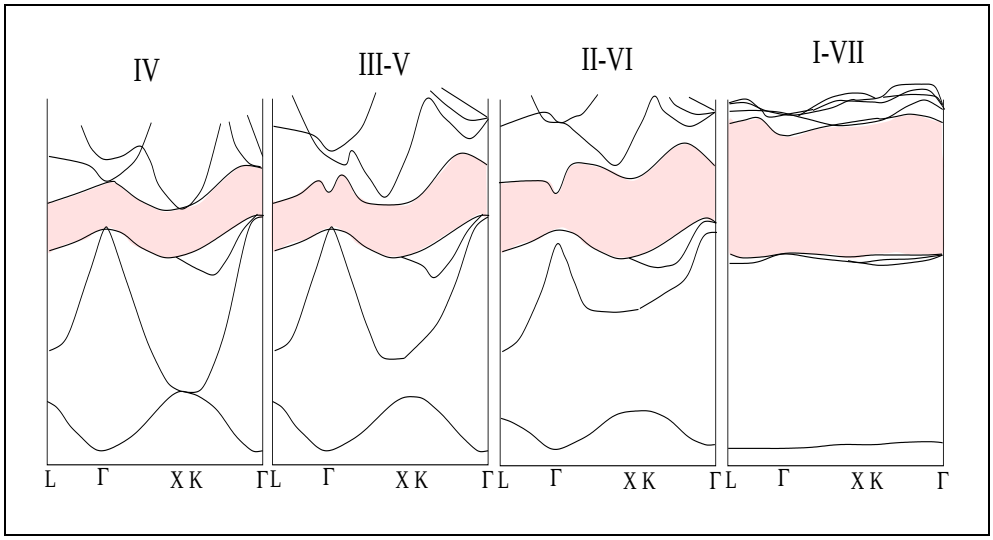

Figure 4.19: Band structures of typical group IV, III-V, II-VI and I-VII solids.

Compounds

The evolution of the band structure from a group IV semiconducting solid to a I-VII ionic crystal is shown in Fig. 4.19. The III-V and II-VI compounds all exhibit tetrahedral bonding (in either the "zinc-sulphide" or "wurzite" crystal structures). In the ionic extreme one finds very narrow bands. In the limit of complete charge transfer the valence charge density resides exclusively on the anion sites. As we go across the series from group IV to I-VII it becomes energetically more favourable to transfer charge and so the band gap widens.

4.7.5 Successes, Failures and Gaps

In this section we have made considerable progress in understanding the diverse nature of solids. Although rather simplistic, the NFEM revealed the origin of the distinction between electrical conductors, semi-conductors and insulators, and a link was established between the occurrence of band gaps and the physics of diffraction.

It would appear that band theory has not taught us anything about the nature of electron scattering mechanisms in metals, nor the anomolous Hall effect measurements we briefly mentioned in the first Problem Sheet.[29] We will try to sort out that stuff in the next few sections where we will take a close look at how electrons move in solids.

As well as explaining the *mechanism* of electron scattering, we must also explain why the scattering rate increases with T. In contrast, the conductivity of semiconductors *increases* with T, but this will be relatively easy to explain. Much more puzzling is the occurrence of *superconductors*, materials which appear to offer no electrical resistance at all.

[29] This is not true, as we shall soon see.

4.8 Electron dynamics: Periodic potentials and the effective mass

In this section we will try to address the question:

How do Bloch electrons move between collisions?

To be more precise we will be interested in their behaviour in the absence and presence of external forces. For now we will forget about electron scattering effects.

The first bit of trickery is to treat the dynamics of the electrons in a solid using the equations of Newtonian dynamics. How can the electron be both a particle (obeying a classical equation of motion) and also a wave (satisfying the Schrödinger wave equation)? We had exactly the same problem with the quantum free electron gas, and the resolution of the dilemma is lurking at the end of Chapter 1 (Sec. 1.2.4). We have to take the Heisenberg uncertainty principle seriously here: if we specify position exactly, then momentum is completely uncertain, and vice versa. The *semi-classical* model of electron dynamics in a solid asserts that reality lies somewhere between these extremes for an electron moving through a solid. For classical dynamics to be valid the electron must be spread over a distance Δx which is very small compared to the length scales of all the processes relevant to its equation of motion (i.e. the mean free path between collisions and the spatial variation of any externally applied fields). For band theory to be meaningful it is necessary for the corresponding uncertainty in wavevector Δk to be much less than the dimensions of the first BZ, or else the energy bands would be smeared beyond recognition. Fortunately, since $k_{BZ} \sim 2\pi/a$ and the collision length is typically $10^2 a$, the uncertainty principle can be satisfied without disturbing either band theory or Newtonian electron dynamics, for most situation sof interest at least.[30]

4.8.1 Electron velocity

Given the comments above it should now be clear why we have been repeatedly referring to the *group velocity* of waves in a crystal. Partial localization in real space means partial delocalization in k-space and so Bloch waves are to be associated with *wavepackets*. We noted earlier that the group velocity of a wavepacket of Bloch waves with wavevector centred on $\underline{k}$ is given by:[31]

$$\underline{v}_g = \frac{1}{\hbar}\nabla_k \epsilon(\underline{k}). \tag{4.46}$$

It is time that we looked at some of the implications of this expression.

1. For free electrons the surfaces of constant ϵ are spheres in k-space. The gradient $\nabla_k \epsilon(\underline{k})$ is therefore parallel to $\underline{k}$. In general, however, the group velocity will

[30] In this paragraph we have been sloppily equating the Bloch wavevector with electron momentum. Very soon

we will see that this really is bad practice, but I am just trying to sketch out the need for a wavepacket approach here, so don't worry too much.

[31] Remember that the t dependence of (stationary) quantum mechanical wavefunctions is in the $\exp(-i\epsilon t/\hbar)$ factor which comes from the time-dependent Schrödinger equation, i.e. we take the angular frequency of the wave as $\omega = \epsilon/\hbar$.

not be parallel to the Bloch wavevector, since constant energy surfaces are not necessarily spherical. This can lead to unexpected dynamical behaviour, very different from that predicted for a free electron, and sometimes counterintuitive.

2. The group velocity of a wavepacket is the velocity at which the wavepacket transports energy through a system. It also gives the *average* velocity of a Bloch electron.[32] Drude would have found this point rather unpalatable. Although the electron velocity varies as it travels across a unit cell in a crystal, its average velocity is not degraded by collisions with the ions, in stark contrast with the Drude-Lorentz or Sommerfeld models. We ended Chapter 1 feeling suspicious of the Drude model of electron scattering in metals. Now we can be quite clear on the matter. The Drude mechanism of electron scattering is wrong - the ion cores in a crystal do not scatter Bloch electrons. Bloch electrons are *stationary states of the Schrödinger equation with the electron-ion interaction fully accounted for.*

3. The non-zero velocity of Bloch electrons also makes a nonesense of the commonly held belief that electrical insulators are the result of electrons being "tied up" in immobile bonds. Bloch electrons travel quite freely through crystals - it is the fact that there is no scope for changing the *distribution* of velocities with an external electrical field that confers insulating properties on a solid. However, inspection of Fig. 4.19 shows that in the ionic 1-7 compounds the energy bands are almost flat. This means electron velocities are approximately zero. This makes perfect sense - we know the electrons in ionic solids are essentially stuck on a specific ion rather than moving freely through the solid. This is not the case for non-ionic insulators such as diamond.

4. The fact that it is $\underline{v}_g$ that describes the motion of the Bloch electron through real space should reassure us of the usefulness of the reduced zone scheme. Although translating an energy band from outside the first zone into it changes the magnitude of $\underline{k}$, it does not change $\nabla_k \epsilon(\underline{k})$.

4.8.2 Force and crystal momentum

Consider applying an external force $\underline{F}^{ext}$ to a Bloch electron. For ease of discussion we can take the specific case of an applied DC electric field $\underline{E}^{ext}$, for which we can write

$$\underline{F}^{ext} = e\underline{E}^{ext} = -e\frac{\partial \phi^{ext}}{\partial \underline{r}} \tag{4.47}$$

where ϕ^{ext} is the electric potential of the external field. However the Bloch electron responds to the field, energy will be conserved:

$$\epsilon(\underline{k}) - e\phi^{ext}(\underline{r}) = \text{constant} \tag{4.48}$$

where the position $\underline{r}$ and wavevector $\underline{k}$ of the electron will change with time t. Differentiating this expression with respect to t yields

$$\underline{F}^{ext} = \hbar\frac{d\underline{k}}{dt}. \tag{4.49}$$

[32] Remember that a Bloch function is not an eigenfunction of the velocity operator $i\hbar\nabla/m$.

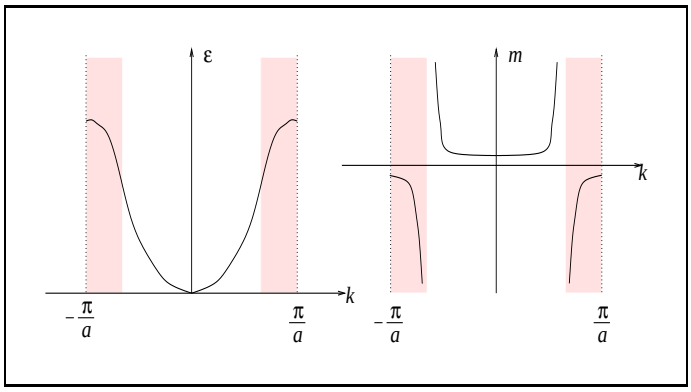

Figure 4.20: Effective mass for the 1D NFE model.

This looks rather like Newton's second law, but we should not be lulled into thinking that the applied force equals the rate of change of the *electron momentum* since this quantity is given by *all* the forces acting on the electron.

On the other hand, we know that Newton's second law must apply somehow. An externally applied force does not act only on a single electron but rather on the crystal as a whole. In the light of Eq. 4.49 it should now seem quite sensible that $\hbar\underline{k}$ is called "crystal momentum".

4.8.3 Electron acceleration and the effective mass

For notational simplicity it is convenient to take the one dimensional case. The electron acceleration caused by an external force is

$$a = \frac{dv_g}{dt} = \frac{1}{\hbar}\frac{dk}{dt}\frac{d^2\epsilon}{dk^2} = F^{ext}\left(\frac{1}{\hbar^2}\frac{d^2\epsilon}{dk^2}\right). \tag{4.50}$$

This equation has the form of Newton's second law $a = F/m^*$ provided we interpret $\hbar^2/(d^2\epsilon/dk^2)$ as an *effective mass* m^*. By relating the external force to the electron acceleration we are again ignoring the force exerted on the electron by the crystal. If we insist on doing this we must fiddle the electron's mass. In other words, a Bloch electron moves under the action of an external force as if it has mass m^* in general different from the true electronic mass so as to account for the effect of the crystal potential.

m^* is shown in Fig. 4.20 for the 1D NFEM. Near $k = 0$, m^* does equal the true electron mass. As we already know, here the bands are free electron like and the electron doesn't notice the crystal potential. We can accept the fact that m^* increases with k since this would appear to reflect the fact that the crystal potential holds onto the electron making it more resistant to acceleration. But the occurrence of *negative effective mass* is harder to stomach at first, since this means that an electron may accelerate in the opposite direction to the applied force! In fact we already know the explanation for this effect. As an external force drives an electron towards a Bragg plane there is an increasing likelihood that it will be Bragg-reflected back in the opposite direction.

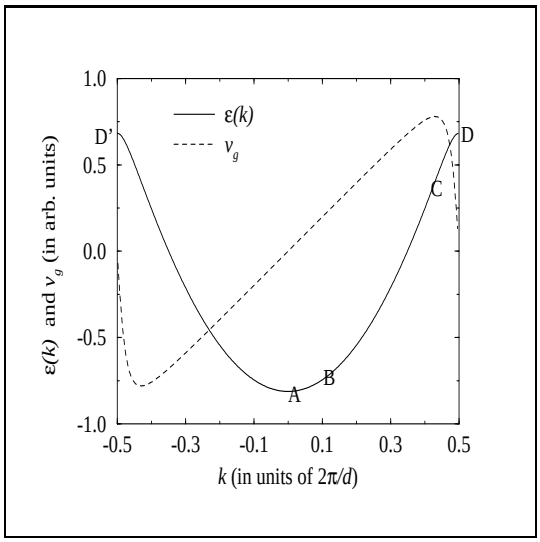

Figure 4.21: The lowest energy band in the 1D NFEM and the corresponding group velocity as a function of $\underline{k}$. The migration of the Bloch electron at $\underline{k} = 0$ upon application of a DC electric field (with the orientation shown) is also shown.

4.9 Electron dynamics: DC Electrical Field and Holes

Now that we have seen the rules of the game, let's continue thinking about the effect of a DC electric field in the 1D NFEM. (But note that we are not talking about electrical conductivity here - that requires both a description of the motion between collisions *and* a description of the collisions. Here we are only talking about the former.)

This section provides a more sophisticated version of our explanation of electrical conduction in Sec. 4.3.3. In a sense we will just be expanding upon and justifying comments made there, but there is a degree of subtlety to be mastered.

4.9.1 A single electron

Consider a single electron in the lowest NFEM band of a crystal. If a DC electric field is applied for a time t then, from Eq. 4.49, it follows that

$$k(t) = k(0) - \frac{eE^{ext}t}{\hbar}. \tag{4.51}$$

Clearly the effect of the electric field is to move the electron along the energy band, as shown in Fig. 4.21. Suppose the electron starts off at position A at the bottom of the band. As the field drives the electron along the energy band from A to B we see a linear increase in the electron velocity. This is free electron behaviour. But as the electron approaches C the acceleration drops and the electron reaches a maximum velocity. With further application of the external force, the electron velocity starts to *decrease*, returning to zero at the

point D at the BZ boundary. This is the negative effective mass in action. At D we know that the Bloch wavefunction is the superposition of plane waves with wavevectors $\pm\pi/a$. This is also true at point D', and so the Bloch waves at D and D' are identical. The electron appears to have left the BZ on the right and re-entered it on the left.

4.9.2 Filled Bands

When more electrons are present in an energy band, they all move along the band like a train. As it crosses the BZ boundary on one side of the zone the train re-emerges on the opposite side, as we saw for the single electron. If the band is completely filled then we can no longer tell if the electrons are shuffling along the band or not. We conclude that *filled bands are inert*. A DC electric field can provoke no response from them.

It may worry you that we are overlooking the possibilty of an electron being promoted to a higher energy band. As we mentioned in Sec. 4.3.3 an extremely strong electric field can do that. In fact *any* field can do it in the free electron limit since successive bands are continuous, as can be seen by comparing Figs. 4.2 and 4.7. Thus *interband* transitions must be accounted for if the field is extraordinarily strong (e.g. lightening) compared to U.

4.9.3 Almost filled bands and holes

We have seen how an almost empty band responds to a DC field and we have seen that a filled (or completely empty) band is inert. It is also important to consider what happens to a band which is almost full, let's say it has space for just one more electron. We can consider such a band to be the sum of a completely filled band together with a single positive charge (to cancel out the extra electron we added). This positive particle is usually called a *hole*. Fig. 4.22 shows that partially filled bands (on the left) can be thought of as a few electrons below the Fermi energy plus a few holes above it instead (on the right).

When the DC field is applied to the nearly filled band the electrons move round the BZ as discussed above and the hole moves in the same direction. Since it has opposite charge to the electron and yet moves in the same direction under the action of an electric field then it must also have the opposite mass to the electron: $m_h^* = -m^*$. In a real solid holes tend to occur at the maxima of energy bands where the curvature $d^2\epsilon/dk^2$ is negative, as shown schematically in Fig. 4.22. This means $m^* < 0$ and so holes tend to have positive mass. Similarly, when bands are only partially filled $d^2\epsilon/dk^2$ is positive and so the few electrons in these states have positive mass. It follows that one can usually describe DC electrical conduction in terms of the transport of two types of carriers: electrons with negative charge and holes with positive charge, but both have positive effective mass.[33] Herein lies the explanation for the strange Hall coefficients of many metals, referred to in Sec. 1.1.4 and Problem Sheet 1 - you have to count the number of electrons and holes since in general $R_H = -1/[e(n_e - n_h)]$.

[33]For any particular partially filled band one has to decide whether to consider it as *either* (i) an empty band plus some electrons, or (ii) a filled band plus some holes. But you don't have to use the same convention for every band.

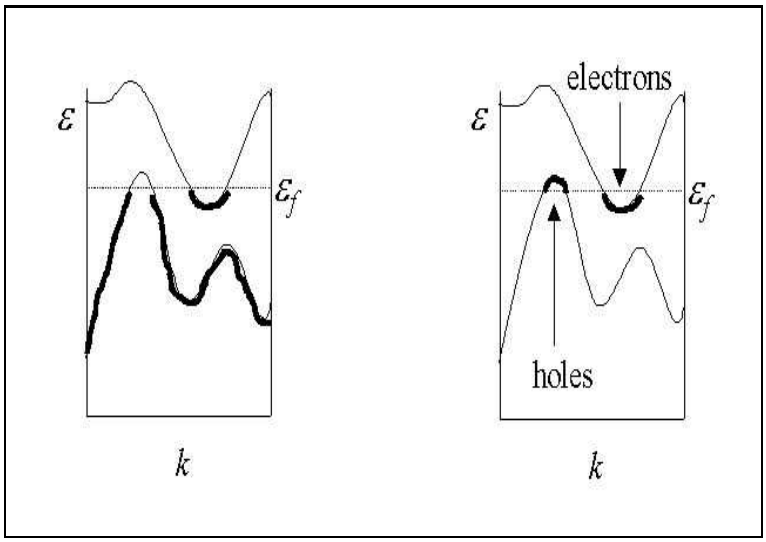

Figure 4.22: Schematic representation of band filling in a solid. The lower band is almost filled while the upper band is almost empty. The electron-hole representation is shown on the right hand side.

4.10 Electron dynamics: Optical properties

Colour can be produced by a variety of physical effects, e.g.
the sky is blue because of the λ^{-4} Rayleigh scattering law,
the sodium lamp is orange because of transitions between its atomic energy levels,
metals are silvery because of the free oscillation of the electron gas,
thin films (such as oil films and soap bubbles) are muliticloured on account of interference,
and so on. There are many solids with charactersitic colours such as copper, gold, ruby, saphire, diamond. Can we explain these?

4.10.1 Metals

The colour of a material is determined by $R(\omega)$, the frequency dependent reflectivity, which in turn is determined by $\sigma(\omega)$, the AC conductivity. The important equations are

$$[n_r(\omega) + i\kappa(\omega)]^2 = \varepsilon(\omega) = 1 + \frac{i\sigma(\omega)}{\omega\varepsilon_0} \tag{4.52}$$

where n_r, κ and ε are the refractive index, extinction coefficient and dielectric function respectively. The reflectivity is related to n_r and κ by the Fresnel equations. For normal incidence we obtain

$$R = \frac{(n_r - 1)^2 + \kappa^2}{(n_r + 1)^2 + \kappa^2}. \tag{4.53}$$

The reflectivity of the free electron gas is shown in Fig. 4.23 where parameters appropriate to metallic Al have been used. The electron gas is highly

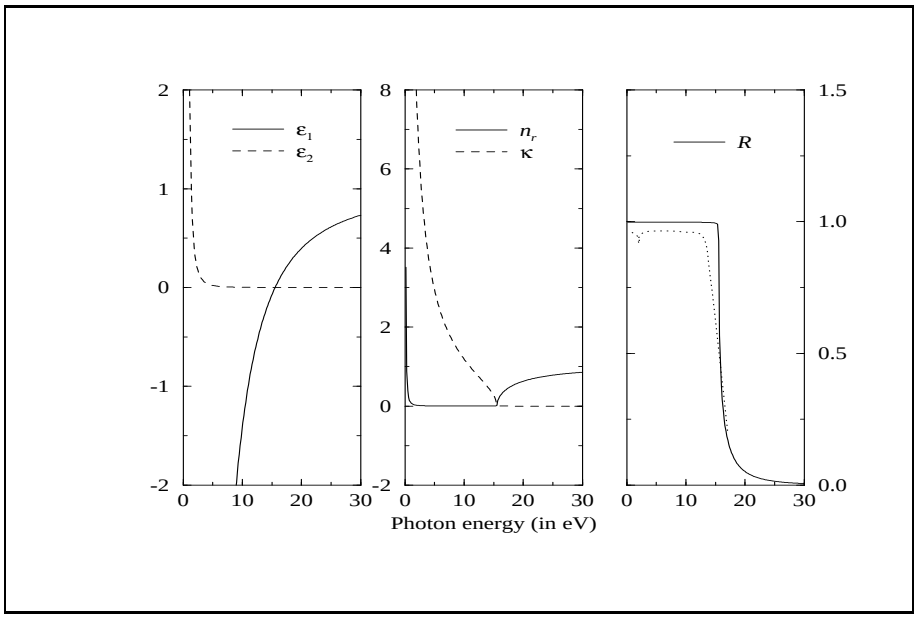

Figure 4.23: The dielectric function $\varepsilon = \varepsilon_1 + i\varepsilon_2$, refractive index n_r, extinction coefficient κ and reflectivity R for the free electron gas (with parameters appropriate for Al). Also shown is the experimental reflectivity of metallic Al (dotted curve).

reflective below a certain frequency, the *plasma frequency*, but is transparent beyond it. The physics here is relatively simple. At low frequency the electric field of an electromagnetic wave causes the Fermi sphere of the electron gas to oscillate back and forth. The energy absorbed by this processes subsequently re-radiated according to Maxwell's equations, so we have (almost) perfect reflection. But if the field tries to drive the electron gas at a frequency which is higher than it is able to respond, then nothing happens. The electron gas becomes transparent. In Problem Sheet 1 we estimated the plasma frequency for the free electron gas, finding it to lie in the UV for typical electron densities found in metals. Fig. 4.23 shows that the free electron model gives a rather good description of the optical properties of Al, but there is the hint of some structure near 2 eV. What else might happen?

We can think of optical radiation as a supply of photons with energy $\hbar\omega$ and momentum $\hbar\underline{q}$. A free electron cannot absorb a photon since the conservation of both energy ($\hbar^2 k'^2/2m = \hbar^2 k^2/2m + \hbar\omega$) and momentum ($\hbar\underline{k}' = \hbar\underline{k} + \hbar\underline{q}$) is not possible.[34] However a Bloch electron can, since the conservation law for *crystal momentum* is

$$\hbar\underline{k}' = \hbar\underline{k} + \hbar\underline{q} + \hbar\underline{K} \tag{4.54}$$

where $\underline{K}$ is a reciprocal lattice vector, as we saw in Sec. 3.5.1. In effect the crystal can recoil to allow the conservation laws to be balanced.

[34]The most a free electron can do is to absorb *some* of the photon's energy. This inelastic scattering of light by free electrons is called *Compton scattering*. This effect leads to a diffuse background of inelastically scattered x-rays in diffraction measurements.

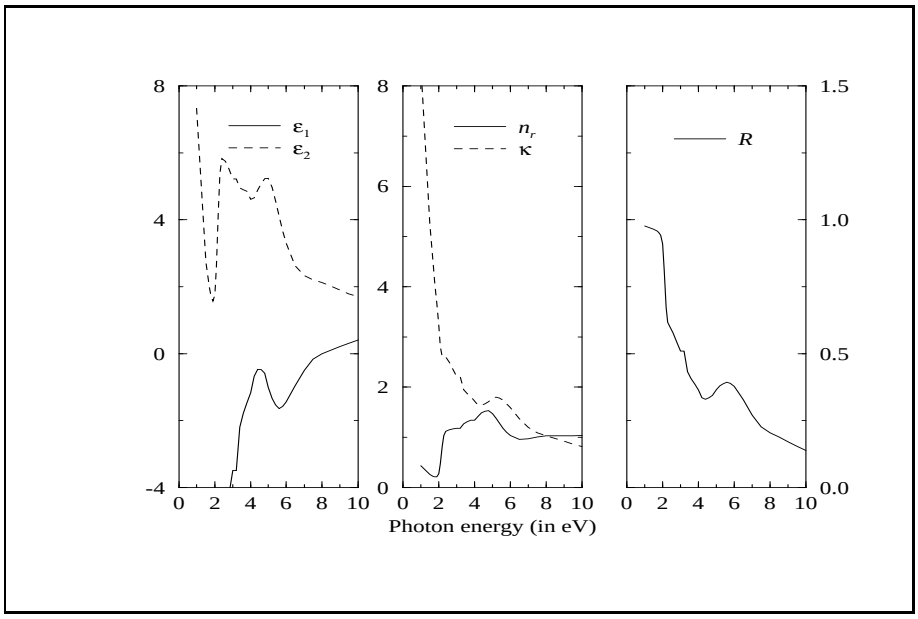

Figure 4.24: The dielectric function $\varepsilon = \varepsilon_1 + i\varepsilon_2$, refractive index n_r, extinction coefficient κ and reflectivity R of metallic Cu.

In the reduced zone scheme we only specify the wavevector of a Bloch wave to within a reciprocal lattice vector, i.e. we plot all the bands in the first BZ. It follows that Eq. 4.54 describes the transition of an electron with wavevector $\underline{k}$ in a particular band to a level with wavevector $\underline{k'} = \underline{k} + \underline{q}$ in a *different band*. These are termed *inter-band transitions*. Typical optical wavelengths are of the order $\sim 10^{-6}$ m giving a momentum $\sim 10^{-27}$ kg m s^{-1}. This is four orders of magnitude smaller than $\hbar\underline{k}_f$ or $\hbar\underline{k}_{BZ}$ and so the change in Bloch wavector when a photon is absorbed is negligible. For this reason the transitions are called *direct* or *vertical*.

Inter-band transitions don't play much of a role in the alkali metals but they are important in the noble metals. The threshold for exciting the $4s$ level of Cu (the long arrow in Fig. **??**) is about 4 eV, but only 2 eV for the d-band. Excitation of the d levels gives rise to strong absorption and the characteristic colour of Cu.[35] The role of shallow valence d levels also accounts for the distinctive colour of Au. The $4d$ band of Ag on the other hand is a bit too far below the Fermi level to produce a colouring in the visible portion of the electromagnetic spectrum. The imaginary part of the complex dielectric function ϵ_2 is closely related to the *absorption coefficient*. The strong absorption between 2 and 5 eV derived from the Cu $3d$ band can clearly be seen in the dielectric function of Cu plotted in Fig. 4.24.

4.10.2 Insulators

We saw in Chapter 3 that solids (except monatomic solids) interact with electromagnetic radiation by creation/destruction of phonons. Experimental inves-

[35] 2 eV is around the red-orange part of the electromagnetic spectrum.

tigation bears out this claim, but since phonon energies are exceedingly small (of order 10^{-3} eV) this mechanism cannot give rise to optical effects in the visible region (where the photon energy is of order 1 eV). Interband transitions are the other chief suspects. However the band gaps of diamond, ruby, saphire for example are several eV giving interband transitions well into the UV. The colours of these gemstones must remain a mystery as far as this course is concerned, but we could at least conclude that the explanation must lie with the breakdown of an assumption made at the very beginning of Chapter 4.

4.10.3 Semiconductors

We haven't got time to say much about the optical properties of semiconductors. We could just return to the point made in Sec. 4.6.7 that the band gap for silicon is 1.1 eV whereas the threshold for interband transitions is 2.5 eV. There is nothing surprising about this - the band gap of sodium is zero but there is a non-zero threshold for interband transitions, as we saw in Problem Sheet 4. But other semiconductors, such as GaAs have the highest occupied state and the lowest empty state at approximately the same position in the BZ. These are *direct* semiconductors. While it is "hard to get light out of silicon", compound semiconductors such as GaAs are extremely useful for making lasers and LEDs.

Semiconductors exhibit *photoconductivity*. They conduct much better when light falls on them we will see why later.

4.11 Collisions revisited

In Sec. 4.4-4.6 we explored the dynamics of electrons in the absence of collisions. This was treated with a subtle fusion of a sophisticated quantum mechanical theory ("band theory") and the more familiar dynamics of Newton. But to explain the conductivity of real metals we return to the Drude collision mathematics (if not Drude's mechanism).

To get some clues as to the origin of electron scattering in metals we can consider some experimental facts.
1. The resistivity of metals increases significantly when they melt.
2. The resistivity of metals increases with the density of structural defects and chemical impurities.
3. The resistivity of metals increases with T. (Linearly at high T?)

4.11.1 Electron-ion scattering

The electron-ion interaction is already included in the Schrödinger equation. Bloch functions are the result of the constructive interference of the electrons in the crystal lattice.

But returning to the approximations at the start of Chapter 4 we see that this argument only holds for a *perfect* crystal. If the crystal melts, contains impurities or has structural defects then electron-ion scattering will increase the electrical resistance. These effects cannot be the whole story however. It is clear from Fig. 4.25 that the dominant effect is T dependent.

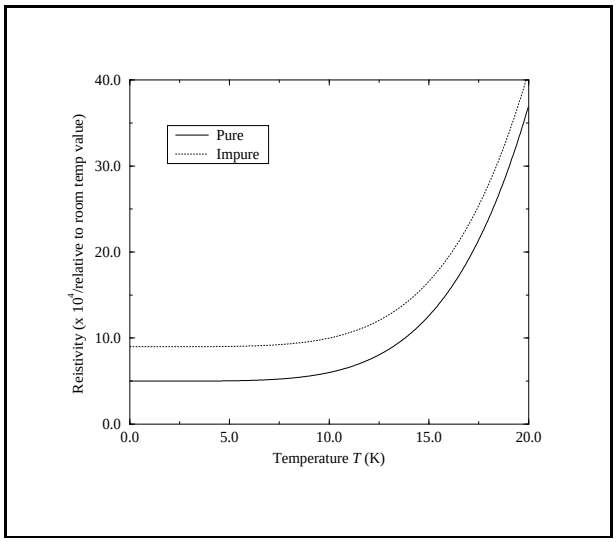

Figure 4.25: Resistivity of pure and impure Na specimens as functions of T.

4.11.2 Electron-electron scattering

The keen observer may have noticed that even the most modern treatments of the electron-electron interactions in solids are fundamentally approximate. We put the electron-electron into the Schrödinger as best we can, but maybe the residual effects we neglect could provide a scattering mechanism. This turns out to be the case (and we will discuss this in SSP) but it's not the dominant effect.

4.11.3 Electron-phonon

Lattice vibrations are to blame, but we will only present circumstantial evidence here. Clearly lattice vibrations perturb the structural perfection of a crystal. The number of phonons incareses linearly with T above the Debye frequency (as we showed in Q5d on Problem Sheet 3) and the resistivity of metals also increases linearly in this regime. Making the case for the defence, you might appeal to Born and Oppenheimer. But the Born-Oppenheimer theorem is only an approximation. It is good for most purposes, but since the atomic and electronic motions in a solid are not rigorously separate phonons do scatter electrons.

We now understand metals pretty well, but there is plenty more physics to be discovered by considering the electrical conductivity of insulators, semiconductors and superconductors. We've run out of time so we'll concentrate on semiconductors on account of their importance in electronics.

4.12 Semiconductors

In contrast to metals, the resistivity of semiconductors decreases with T and they are said to have a *negative coefficient of resistance*. It was this anomalous behaviour that got nineteenth century physicists such as Faraday interested in

semiconductors. By the turn of the century semiconductors had been found to exhibit photoconductivity (an increase in conductivity when light when a solid is exposed to light), an extremely large Seebeck effect (an electric field generated by a thermal gradient), strange rectifying effects at the junctions of dissimilar semiconductors, and strange Hall coefficients. The experimental study of these materials was initially rather confusing, with inconsistent results frequently obtained by different researchers. We will be able to explain this, but first let's think about the basics.

4.12.1 Charge carriers in semiconductors

At $T = 0$ semiconductors have no partially filled bands, but for $T > 0$ there is a finite probability that electrons can be thermally excited across the band gap.[36] It is therefore much more convenient to treat semiconductors using the electron-hole formalism illustrated on the right hand side of Fig. 4.15. For semiconductors, electrons can only exist above the Fermi level while holes can only exist below it. Of course the extent to which thermal excitation can occur is determined by the energy gap between the "valence band" and the "conduction band".[37] The electronic properties of semiconductors are determined by these few electrons promoted to the lowest conduction band and the holes left behind at the top of the valence band. In the region of the band extrema we can approximate the bands by parabolae.

It can be shown that in this limit the conductivity formula reduces to Drude-like form, provided we account for both conduction by holes and electrons:

$$\sigma = e^2\tau\left(\frac{n}{m_n^*} + \frac{p}{m_p^*}\right) \tag{4.55}$$

where n and p are the number of conduction electrons and valence holes per unit volume, and the m^* are the appropriate effective masses. This equation is often rewritten

$$\sigma = ne\mu_n + pe\mu_p \tag{4.56}$$

since this makes a separation between the *quantity* of carriers and their *mobility* μ given by $e\tau/m^*$. The full band structure of Si is shown in Fig. 4.19 on page 125. Notice that the valence band maximum occurs at the Γ point where bands with different curvature meet, so we have some "heavy" and some "light" holes which therefore have different mobilities.

The carrier densities n and p are strongly dependent on temperature. As we saw in Chapter 1, these quantities can be expressed as integrals:

$$n = \int_{\epsilon_{CB}}^{\infty} g_{CB}^{\epsilon}\, f^{\epsilon}(T)\, d\epsilon \qquad p = \int_{-\infty}^{\epsilon_{VB}} g_{VB}^{\epsilon}\, [1 - f^{\epsilon}(T)]\, d\epsilon. \tag{4.57}$$

[36] At the start of Chapter 4 we stated that only the $T = 0$ limit is of any interest when it comes to electronic structure. For metals the Fermi temperature is so high that room temperature, for example, may as well be considered as $T = 0$. But T is of the utmost importance in semiconductor physics.

[37] We have previously used the term *valence levels* to mean those levels derived from the outermost electronic shells of atoms in a solid. The term *conduction electron* in the context of metals referred to those free electrons giving rise to the conduction of electricity. In insulators we adopt a more careful terminology, reserving the term *valence level* for those states which are filled at $T = 0$, and the term *conduction level* for those states which are unoccupied at $T = 0$. In addition the term "valence band" in semiconductor physics is taken to mean all the valence energy bands, and similarly for the term "conduction band".

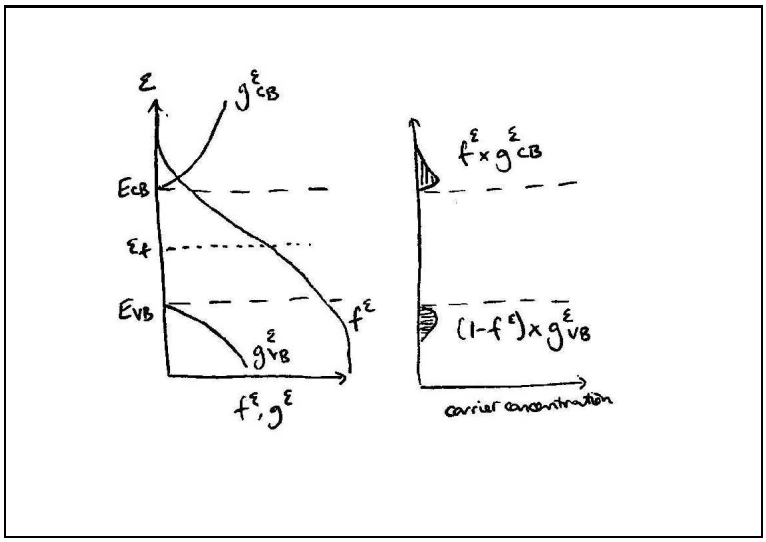

Figure 4.26: Fermi-Dirac function and density of energy levels in a semiconductor. The Fermi level is determined by the requirement $n = p$.

In the parabolic approximation, the densities of energy levels are given by:[38].

$$g^{\epsilon}_{CB} = \frac{1}{2\pi^2}\left(\frac{2m^*_n}{\hbar^2}\right)^{3/2}\sqrt{\epsilon - \epsilon_{CB}}, \qquad \epsilon > \epsilon_{CB} \tag{4.58}$$

$$g^{\epsilon}_{VB} = \frac{1}{2\pi^2}\left(\frac{2m^*_p}{\hbar^2}\right)^{3/2}\sqrt{\epsilon - \epsilon_{VB}}, \qquad \epsilon < \epsilon_{VB} \tag{4.59}$$

The temperature dependence of n and p is derived from the occupation factor f. This is just the Fermi-Dirac distribution, but a problem with this is: what do we take as the Fermi level in a semiconductor? We know that the VB holes in a (pure) semiconductor are created when electrons are promoted to the CB, and so n and p must be equal. This condition determines ϵ_f as illustrated in Fig. 4.26.

Since the Fermi energy of a (pure) semiconductor falls roughly in the middle of the band gap, it follows that the occupation factor f is rather small. In fact it can easily be shown that since $\epsilon - \epsilon_f \gg k_bT$

$$\frac{1}{\exp[(\epsilon - \epsilon_f)/(k_bT)] + 1} \approx \exp\left(-\frac{\epsilon - \epsilon_f}{k_bT}\right) \ll 1 \tag{4.60}$$

which means the Fermi-Dirac distribution approaches the Maxwell-Boltzmann one. Physically this just means that the CB-electron and VB-hole occupation probabilities are so low in a semiconductor that there is no danger of ever violating the Pauli principle. It is then easy to show that

$$n = 2\left(\frac{2\pi m^*_n k_bT}{\hbar^2}\right)^{3/2}\exp\left(-\frac{\epsilon_{CB} - \epsilon_f}{k_bT}\right) \qquad p = 2\left(\frac{2\pi m^*_p k_bT}{\hbar^2}\right)^{3/2}\exp\left(\frac{\epsilon_{VB} - \epsilon_f}{k_bT}\right). \tag{4.61}$$

[38]From Chapter 1, $\epsilon \propto k^2 \Rightarrow g^{\epsilon} \propto \sqrt{\epsilon}$

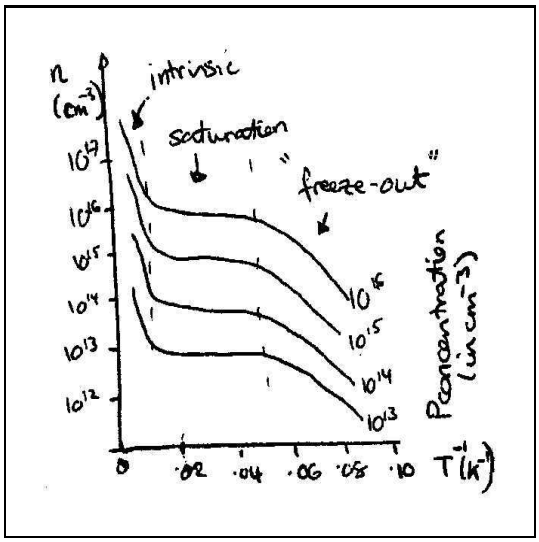

Figure 4.27: n as function of reciprocal temperature for P doped Ge.

A loose end which we can dispense with now is the observation of photoconductivity in semiconductors (an increase in conductivity upon exposure to light). Band theory simply explains this in terms of the creation of charge carriers, i.e. VB holes and CB electrons, upon absorption of light. This effect is quite useful, for example in making sensors etc.

4.12.2 Dopants

The typical room temperature carrier densities obtained by the above equations are $\sim 10^{16}$ m^{-3}, which is too small for practical semiconductor devices to operate. We can increase the electron/hole concentrations by adding small quantities of pentavalent/trivalent atoms to a Si crystal. Such *dopants* are referred to as *donors* and *acceptors* respectively. A semiconductor with a net excess of electrons is referred to as *n-type* while that with an excess of holes is called *p-type*. To understand the effect of dopants, let's consider what happens to a P impurity in Si. The P atom takes the place of a host atom, using four of its valence electrons to bond to four neighbouring Si atoms. The remaining electron is relatively loosely bound. It's not totally free since the extra nuclear charge of the P atom keeps hold of it. We can estimate just how bound the donor electron is by regarding the P impurity as a hydrogen atom embedded in a dielectric medium. It follows that the ionization energy of the extra electron are (in eV)

$$I = \frac{13.6}{i^2}\frac{m_n^*}{m}\frac{1}{\kappa_{\rm Si}^2} \tag{4.62}$$

where m_n^* is the effective mass of a CB electron in Si, m is the true electron mass, $\kappa_{\rm Si}$ is the dielectric constant of Si, and i is the principal quantum number. This shows that P impurities create donor levels which are only ~ 0.05 eV below ϵ_{CB}. Occupation of these donor levels leads to movement of the Fermi level. In n-type semiconductors it moves up towards the CB edge, while in p-type material it moves down towards the VB edge. This simple model works surprisingly well.

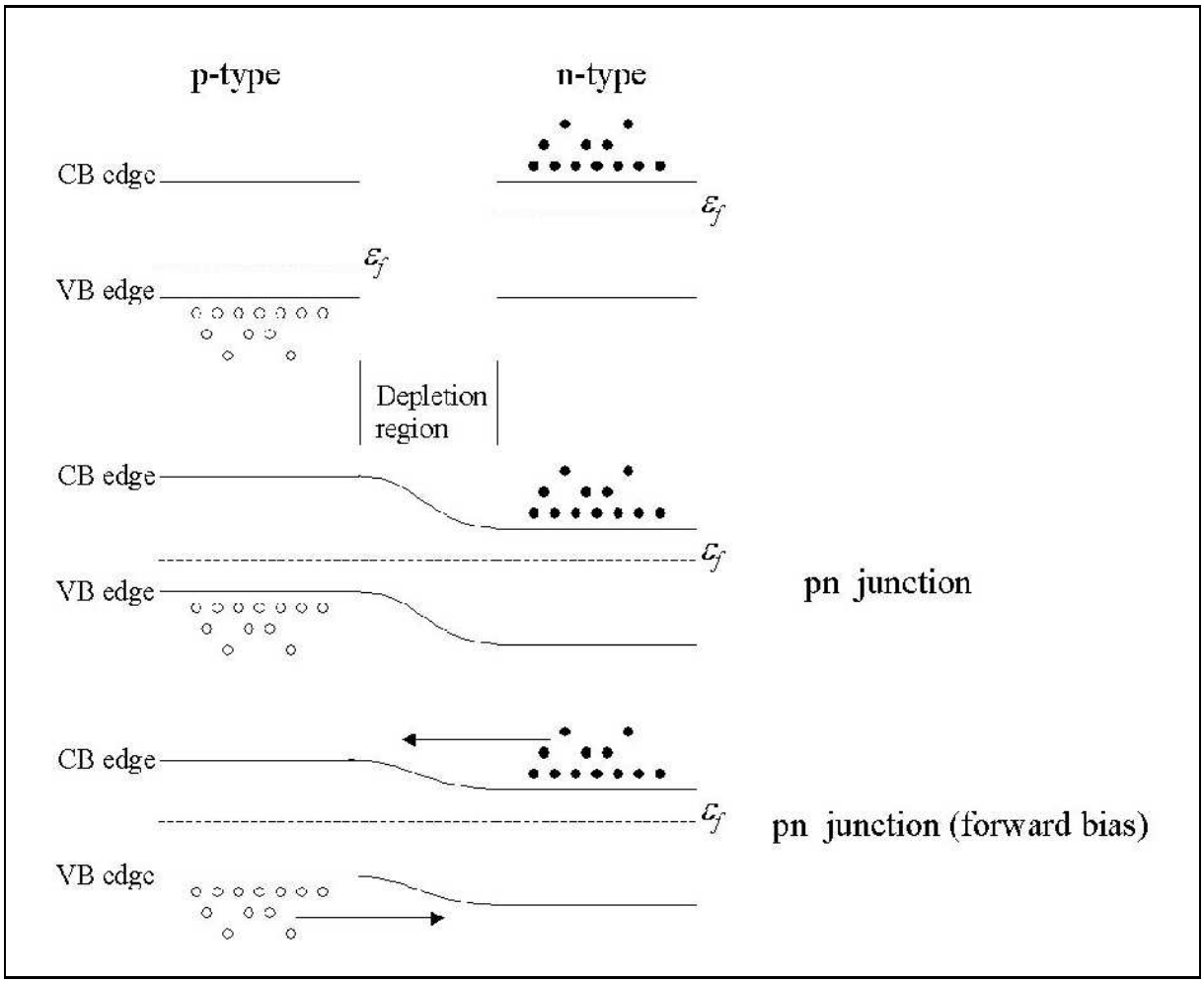

Figure 4.28: Energy levels of p and n-type semiconductor and the pn junction. Also shown is the pn junction under forward bias.

We can now see why the early measurements on semiconductors were plagued by inconsistencies - the conductivity of such materials is inordinately sensitive to minute quantities of impurites. The carrier density of Ge doped with P is shown in Fig. 4.27. Three distinct regions are observed. At low T only some of the donor electrons are ionized and this is called the *freeze-out* regime. On increasing T increasing numbers and eventually all of them are ionized - the *saturation* regime. Further increase in T leads to an increase in n due to thermal excitation of electrons from the VB - the intrinsic regime.

4.12.3 Semiconductor devices

Here we have time to mention only the most simple semiconductor devices, but given their monumental impact on all forms of human activity it is important to at least make a start.

The pn junction

Consider first two dissimilar metals. In general these will have different Fermi energies, so let's suppose that metal A has a higher Fermi energy than metal B. When we place samples of A and B in contact there will be electrons in occupied states on metal A which have a higher energy than some of the empty states on metal B. Electrons will then flow from A to B until the Fermi level is uniform across the boundary. This charge redistribution creates a potential difference at the junction of the two metals called the *contact potential*.

Now let's consider two dissimilar semiconductors in contact, let's say p-type Si in contact with n-type Si, as shown in Fig. 4.28. Electrons from the n-type side and holes from the p-type side diffuse across the boundary so as to balance the Fermi levels either side of the boundary. But the electrons and holes in the vicinity of the boundary can annihilate each other (known as "recombination") leading to a net reduction in charge carriers in this region which is therefore known as the *depletion region*. In the depletion region the excess nuclear charges remain, and we can think of the potential step they cause across the boundary to be analogous to the contact potential at the boundary two metals.

When there is no externally applied field, electrons prefer to reside in the CB on the n-type side of the junction since here their potential energy is lower than it would be on the p-type side. However, if we apply a battery to the junction such that the positive terminal is connected to the p-type side of the junction (known as "forward bias") then electrons on the n-type side start to become tempted to move across, as shown in Fig. 4.28, and so a current can flow. Alternatively, reverse biasing the pn junction makes this process less tempting and so no current flows. A similar discussion can be made for the movement of holes. We should note that we have only been concentrating on the majority carriers - ideally we should also recognise that for $T > 0$ there are a few electrons in the CB of p-type Si and also some holes in the VB on n-type Si. Nonetheless our simplified treatment has allowed us to see that the pn juction acts as a "rectifier" or "diode": it only allows current to pass in one direction.

The transistor

Transistors can are made by suitably joining two pn junctions back to back. We haven't got time to go discuss their operation (see "Semiconductor Physics" next term) but we can note that transistors can be used as current amplifiers or as electronic switches. This later application allows electronic digital computation.

The transistor must rank as one of the most significant inventions of the twentieth century, but its vast impact on society was not really felt until other technology, molecular beam epitaxy (MBE) and related techniques, was developed. MBE allows semiconductor devices to be constructed one atomic layer at a time, allowing extraordinary control and purity. Once techniques were developed for patterning surfaces (e.g. to etch away areas, fill them with metal to make tiny wires, or with insulators etc.) on a microscopic scale it became possible to imagine single integrated devices each containing vast numbers of individual components. VLSI (very large scale integration) as much as the transistor itself has given birth to the IT revolution.

We live in interesting times from this point of view, as current state of the art microelectronic devices use components at the nanoscale. We have been used to

exponential growth of computing power (known as "Moore's Law"), but there is little scope for further shrinkage of conventional solid state electronics. So what will come next?

4.13 Insulators

In Problem sheet 4 we found that the observed conductivity of insulators cannot be explained by the thermal excitation mechanism of semiconductors. This, along with the optical properties of insulators, will remain a mystery.

4.14 Superconductivity

Superconductivity is spectacular. The characteristic properties of this class of material are:
(i) Lack of DC electrical resistance. It is not small, it is completely absent.
(ii) Perfect diamagnetism, although we have discussed what this is.
(iii) An apparent energy gap at the Fermi level.
This is quite spectacular, and the explanation is extremely subtle and highly specialised, but superconductors are not rare. A large number of the elements display superconducting states.

It should be clear, even from the meagre information given above, that superconductivity cannot be explained by refining our understanding of metals. Rather a complete revolution is needed. It turns out that superconductivity relies on electrons being bound in pairs. This seems somewhat unlikely. It is also fundamentally beyond the scope of band theory which relies on the one-electron approximation. We can't go into this here.

4.15 Summary

In this Chapter we have assumed that
(i) we know where the atoms in a particular crystal are located (i.e. we know the crystal structure), and
(ii) the atoms are stationary.
We would like (eventually) to offer a more fundamental treatment of solids but here we built up a theory of electronic structure only. Our approach was based on the One Electron Approximation which allows us to treat only one electron at a time, assuming it moves in some (periodic) potential U.

The wavefunction of an electron moving in a periodic potential must have the form given by Bloch's theorem: $\psi_{\underline{k}}(\underline{r}) = u_{\underline{k}}(\underline{r}) \exp(i\underline{k}.\underline{r})$, where u has the periodicity of the direct lattice.[39] Application of periodic boundary conditions shows that the density of allowed allowed $\underline{k}$ vectors is exactly the same as in the free electron model, but since Bloch waves are not momentum eigenstates it is not yet clear how the Bloch wavevector $\underline{k}$ relates to kinematic momentum. We showed that for any Bloch wave with wavevector $\underline{k}_1$ outside the first BZ it is always possible to find an equivalent one with wavevector $\underline{k}_2$ (where $\underline{k}_1$ and

[39] Bloch's theorem must be true for both core and valence electrons but we are particularly interested in the latter since these are responsible for the bonds which give rise to the solid state.

$\underline{k}_2$ differ by a reciprocal lattice vector) inside the first zone. Thus we adopt the reduced zone scheme - it is only necessary to consider $\underline{k}$ to lie in the first BZ.

Expanding the Bloch wavefunctions and the crystal potential using Fourier series we get the "k-space Schrödinger equation" which can be viewed as a matrix equation which we must solve at each k-point. The thing to notice here is that we must truncate the Fourier series of $u_{\underline{k}}(\underline{r})$ in Eq. 1.11 and $U(\underline{r})$ in Eq. 1.18 so that the matrix equation has finite size (i.e. a finite number of rows and columns). If we are to truncate the Fourier series of U and u then it is essential that they converge rapidly in reciprocal space. It is not at all clear that this will be the case since the Coulombic potential between an ion core and a valence electron decays only as $1/k^2$ in reciprocal space. In fact the problem is much worse than we may have supposed since orthogonalising the valence levels to the core levels requires the wavefunctions of the former to be extremely rapidly varying in the core region.

In effect the orthogonality constraint (the Pauli principle), by expelling the valence electrons from the core region of an atom, tends to partially cancel the strong Coulombic attraction between valence electron and core. Consideration of the importance of this effect for particular atoms leads us to an understanding of the chemical trends we observe in the Periodic Table of elements, and in particular why some elements give rise to free-electron-like metallic states in which the valence electrons appear to feel no attraction for the ion cores at all. If we limit ourselves to finding the valence electron wavefunctions outside of the core region then we can adopt the pseudopotential method in which the core electrons (and hence the orthogonality problem) are totally removed from the calculation. Instead we assume the valence electrons move subject to a weak potential, the pseudopotential . This is constructed such that the wavefunction inside the core region is wrong *but smooth*, and is correct outside the core.[40]

The potential U comprises not only $\sum U^{PS}$ but also a term arising from electron-electron interactions. The repulsive energy between an electron and the average charge density produced by all the others is called the Hartree energy. This term includes electron-electron repulsions in an average way but the repulsion an electron really feels at any particular instant is determined by the positions of all the others at that instant. To be technical, we have assumed that electron motions are *not correlated*. We know that this cannot be correct because of (i) antisymmetry (the "exchange effect"), and (ii) Coulomb correlations, but the assumption is highly convenient since it allows us to attack the Schrödinger equation just one electron at a time. Even this is not as simple as it sounds since the Hartree potential depends on the wavefunctions of all the electrons in the solid, which one does not know at the outset, so we must adopt an iterative approach in order to achieve self-consistency.

One can go beyond the Hartree approximation by trying to patch in exchange and correlation terms to the one-electron Schrödinger equation. This can be done exactly for the exchange term to give the Hartree-Fock equations and these can be solved for the homogeneous electron gas. The exchange effect gives rise to a negative contribution to the total energy of the system (sometimes called "jellium"), suggesting the origin of cohesion in metals. It is important to notice that the HF electron gas is homogeneous *on average*, but at any instant

[40]Since the pseudopotential is determined by the nucleus and core electrons it can be found for a particular element by performing calculations for a single atom for which the answers are known.

there is a region around each electron in which other electrons with the same spin alignment are forbidden - the *exchange hole*. The HF model leads to the electronic states being spread over a much wider energy range than predicted by the free-electron model (which neglects all electron-electron and electron-ion effects, or assumes that they cancel each other out).

There are some other strange consequences of the HF electron gas model which render it quite unpallatable. The lesson here is that it is not always worth putting in one effect exactly if one neglects other effects which may be equally important. The way forward here is to approximate both exchange and correlation together. In fact both can be treated exactly for the homogeneous electron gas, prompting the local density approximation to density functional theory, as discussed in the previous section.

We now have a scheme for computing the electronic structure of crystals from the Schrödinger equation (see Fig. 4.17). Before we elaborate this framework further we need to see if it actually works. Since we have made a number of crucial approximations and assumptions it is also important to convince ourselves that the scheme *should* work. Since electron-electron interactions are strong and their motions correlated one might not expect the one-electron approximation to work very well. We can perhaps anticipate an explanation: by including an exchange-correlation potential in the one-electron Schrödinger equation we are effectively replacing the strongly interacting and correlated electron system with weakly interacting uncorrelated quasiparticles.

Chapter 5

Bringing it all together

So far we covered essentially three topics: crystal structure, atomic dynamics, and electronic structure. One may wonder how the three main strands of this branch of physics fit together. We studied structure in considerable depth, discussing the types of structure which form, how one describes then in the most compact way, how one could experimentally determine the structures of unknown materials, and we tried to discuss in a qualitative way why a particular material adopts a specific structure. We discussed the vibrational properties of crystals, with the primary motivation of explaining the thermal properties of crystals. This we achieved using a "ball and spring" model in the belief that in some vague way the chemical bonds in a solid constitute the springs. We then contemplated the behaviour of electrons in a crystal, but notice that *the atoms were assumed stationary in some predetermined structure*. We discovered an explanation for the occurrence of metals insulators and semiconductors in nature and with the pseudopotential idea we tried to explain trends observed in the Periodic Table.

It is clear that the three main concepts mentioned above must be related. Electrons are the glue binding atoms together. The electronic structure of a material must therefore play a major role in determining its total energy and hence the crystal structure it favours, the stability of that structure, as well as the "spring constants" which determine the atomic dynamics. We aim now to arrive at a unified theoretical formalism that treats the three strands of solid state physics in a coherent and self-contained manner. In this way we will not only arrive at an appreciation of how it all fits together, we will also assemble all the knowledge and techniques necessary for computing the properties of a material the Schrödinger equation.

That this can be done is quite remarkable. We will have the ability to predict, for example, that silicon should adopt the diamond crystal structure with an interatomic bond length of 2.35 Å. Experimental observation shows that it is indeed so, giving us confidence to use the same methods to predict what will happen to silicon at extra-ordinarily high pressure or temperature where doing the experimental measurements may be either hazardous, expensive, time-consuming or even impossible. Similarly we will have the power to invent new materials and explore their properties, eliminating useless ones without wasting time fabricating and testing them in real life.

Physicists are very good at devising clever and robust theoretical frameworks

such as this, but we are not satisfied with just churning numbers out with a computer. We want to be able to *understand* why the numbers come out as they do. Can we look a little deeper and appreciate *why* silicon adopts the diamond structure while lead (in the same chemical group) does not? Why do the metals Mn, Fe, Co, Ni (adjacent in the Periodic Table) each have a different crystal structure? We need some qualitative insight to go with our computing power. Indeed it is worth pointing out that the methods we have at our disposal for the simulation of material properties seem to work far better than we have a right to expect. Explaining why these methods work at all is therefore a major challenge.

It should be mentioned that our powers are limited and some things just come out wrong. We are defeated for a variety of reasons. Certain mechanical properties of solids deviate enormously from our predictions because of the existence in the real world of defects. Such phenomena are not necessarily the result of sloppiness on the part of the scientist preparing samples for study. Rather their existence is dictated by the laws of nature (thermodynamics in fact). It is not easy to come to terms with this since our theoretical results and computer programs tend to work best for perfect crystals where every unit cell is rigorously identical.

There are fundamental obstacles of a different kind. Quite simply our assumptions and approximations fail in grand style on occasions. Thus modern solid state physics is a mixture of experimental investigations, quite difficult theoretical ideas, state-of-the-art numerical simulation techniques, and simple models that help us understand what it all means.

5.1 Cohesion: Qualitative Observations

To predict the crystal structure of a solid we need to calculate the structure for which E_{gs}^{tot}, the ground state total energy of the solid, is the lowest. To quantify the strength with which atoms are bonded together in a solid we define the cohesive energy:

$$E^{coh} = \frac{1}{N}\left(E_{\infty}^{tot} - E_{gs}^{tot}\right) \tag{5.1}$$

where E_{∞}^{tot} is the total energy when the solid is disassembled into its N constituents. In words this equation says:

• *The cohesive energy is the energy required to disassemble a solid into its constituent parts divided by the number of constituents.*

For different classes of solid it is appropriate to consider different types of constituent.

5.1.1 Homo-nuclear systems

Solids comprising a single type of atom must adopt either the covalent, metallic or Van der Waals (also known as "molecular") bonding strategy. We saw in §2.1.2 that the covalent concept is based upon the formation of so-called "bonding" and "anti-bonding" molecular orbitals, the former having lower energy than the corresponding atomic orbitals while the latter have higher energy. These

energy differences were quantified by the "hopping integral", h, defined in Eq. 2.9. This equation shows that the deeper the energy wells felt by the valence electrons (see Fig. 4.6), the greater the magnitude of h, The covalent strategy works in molecules when there are more electrons in the bonding than the anti-bonding states, and the same argument applies in extended solids (except that the molecular orbitals are broadened into energy bands).

In §4.6.3 we discussed briefly an insightful approach to understanding trends in the strength of these atomic potential wells. We noted that in principle H, Li, Na, K, Rb and Cs can each be viewed as a positive ion surrounded by a single valence electron, however the ionization potential (the energy required to rip off the valence electron) decreases significantly down this group, as if the potential felt by the electron is getting weaker. We explained this trend by recalling that the valence electron in a Cs atom, for example, resides in the 6s shell and its wavefunction must be orthogonal to the 1s, 2s, 3s, 4s and 5s wavefunctions.[1] To achieve orthogonality (i.e. $\int \psi^*_{ns}\psi_{ms}dr = \delta_{mn}$) to these inner shells the 6s wavefunction contains 5 radial nodes (where the radial wavefunction has zero amplitude) in the inner region of the atom. Since the wavefunction is "wiggly" in the core region the electron travels very quickly[2] there. In effect the valence electron of Cs is prevented from lingering in the region where the Coulomb attractive is strong. It is as if it cannot "feel" the nucleus. This is why free electron type behaviour decreases down the Periodic Table. Moving across the Periodic Table on the other hand increases the nuclear charge but brings no additional orthogonality requirements, thus U^{ps} increases across the Periodic Table.

These considerations suggest that metals should be formed by elements to the lower left of the Periodic Table. Elements towards the upper right will tend to have strong U^{ps} and will occupy fewer anti-bonding than bonding states, so covalency rules. Elements to the extreme right (the "noble gases") exhibit strong U^{ps} but covalency is not favoured since as many electrons will be in anti-bonding as bonding states. Only the Van der Waals bonding mechanism remains. As we saw in §2, the potential energy for two molecules is given by the Lennard-Jones potential [3]

$$V(r) = 4V_0\left[\left(\frac{\ell}{r}\right)^{12} - \left(\frac{\ell}{r}\right)^{6}\right]. \tag{5.2}$$

The constants ℓ and V_0 represent the range and strength of the potential. To calculate E^{coh} we must consider the mutual interactions of all molecules (in this case atoms). Let's take the origin to be the position of molecule 1. The energy of interaction between the molecule at the origin and all the others in the solid is then the sum of $V_i(r_i)$, where i labels all the other molecules and r_i are their distances from the origin. To get the total energy of a molecular solid we need to multiply this quantity by N, the total number of molecules, and divide by 2,

[1] It is automatically orthogonal to all p and d shells on account of the angular part of the wavefunction.

[2] The velocity operator is proportional to d/dr.

[3] Since the molecules in a van der Waals solid are only weakly perturbed, we can get a rough understanding of cohesion by setting the *intra*-molecular energy (i.e. E^{tot}_{∞} in Eq. 5.1) to zero and equating the total energy of the solid by the *inter*-molecular interaction energy.

to avoid double counting. The cohesive energy is therefore

$$E^{coh} = -\frac{1}{2}\sum_{i\neq 1} V_i(r_i) \tag{5.3}$$

which can be written in the form

$$E^{coh} = -2V_0\left[A\left(\frac{\ell}{a}\right)^{12} - C\left(\frac{\ell}{a}\right)^{6}\right]. \tag{5.4}$$

where a is the nearest neighbour distance and A and C are structure dependent constants defined by

$$A = \sum_{i\neq 1}\left(\frac{a}{r_i}\right)^{12}, \quad C = \sum_{i\neq 1}\left(\frac{a}{r_i}\right)^{6}. \tag{5.5}$$

Since van der Waals interactions are short range, only the first few nearest neighbours contribute to the sums defining A and C. Evaluating these quantities reveals that van der Vaals bonding favours close packed structures, as we may have guessed.

At this point we should consider whether metals should form at all. In §2.1.5 we noted that delocalization allows a lowering of electron kinetic energy. However this is a positive energy, whereas an electron on an atom has negative energy (i.e. is bound). Why do the atoms stick together? Part of the answer was revealed in §4.7.2. There we briefly considered the Hartree-Fock treatment of a homogeneous electron gas, sometimes referred to as "jellium". In jellium the positive charge of the ions is assumed to be smeared out uniformly. The electron charge density must therefore also be uniform since there is no reason for it to be otherwise. If the electronic motions are assumed to be uncorrelated (i.e. the Hartree model is assumed) the only contributions to the total energy are the KE of the electrons (the positive charge is assumed immobile), the e-e repulsion, the ion-ion repulsion and the e-ion attraction. For uniform distributions the electrostatic terms cancel leaving only the electron KE. The Hartree-Fock model adds the constraint of Fermi-Dirac statistics (i.e. the wavefunction of the electron system must be anti-symmetric) leading to a reduction in the e-e repulsion. Effectively electrons with parallel spin repel each other less than Hartree assumed by because they are forbidden from getting close to each other. The HF energy of jellium was given in Eq. 4.40. Accounting also for Coulomb correlation as well as the exchange effect, it can be shown that the total energy of jellium is (in eV per electron)

$$E^{tot} = \frac{30.1}{(r_s/a_0)^2} - \frac{12.5}{(r_s/a_0)} - (1.56 - 0.42\ln(r_s/a_0)) \tag{5.6}$$

where the three terms on the right represent kinetic energy, exchange energy and correlation energy respectively. a_0 is the Bohr radius (0.53 Å) and we are again using the variable r_s which is just the radius of the sphere whose volume is equal to the volume per conduction electron in the metal. r_s is related to the electron density n by the relation

$$\frac{1}{n} = \frac{4}{3}\pi r_s^3. \tag{5.7}$$

We can see that jellium is stabilised by the exchange-correlation terms, the minimum energy per electron of -2.2 eV occurring for $r_s = 2.2$ Å. The "exchange-correlation hole" surrounding each electron allows it to enjoy a net attraction from the underlying positive charge.

There are a couple of problems with this:

(i) Values of r_s for real free-electron-like metals lie in the range 1 to 3 Å. The value predicted by the jellium calculation is therefore quite sensible, but what makes r_s for one metal quite different to that for another?
(ii) Experiment tells us that it takes ~ 1 eV per atom to pull metals like sodium apart into individual atoms (i.e. $E^{coh} \sim 1$ eV). But the sodium atom still has some energy, just as the energy of the hydrogen atom is -13.6 eV. For sodium this number, the negative of the ionisation potential, is ~ -5 eV. Clearly if $E^{tot}_{\infty} \sim -5$ eV and $E^{coh} \sim 1$ eV then it follows from Eq. 5.1 that E^{tot}_{gs} should be about -6 eV. The jellium model is missing an important contribution to the total energy of real metals.

You can probably guess the problem - we need to do something more realistic than smearing the positive charge out uniformly.

The simplest way of doing this is to replace the smeared out positive jellium charge with ions embedded in the electron gas at the correct places. In a metal, each Wigner-Seitz cell will be neutral and so we need only consider the interactions within the WS cell. We can estimate the energetics by making a few simple and quite plausible assumptions:
(i) We assume that the electron density is approximately homogeneous.
(ii) We replace the true WS cell (a polygonal volume) with a sphere (of radius R_{WS} which has the same volume.
(iii) At each ion position we place the appropriate pseudopotential .
(iv) We assume for simplicity that there is only one atom per unit cell.
The total energy of the metal (in eV per atom) is then

$$E^{tot} = Z\left\{\frac{30.1}{(r_s/a_0)^2} - \frac{12.5}{(r_s/a_0)} - (1.56 - 0.42\ln(r_s/a_0))\right\}$$

$$- \int_{WS} \rho(\underline{r})\, U^{ps}(\underline{r})\, d\underline{r} + \frac{1}{2}\frac{e^2}{4\pi\varepsilon_0}\int_{WS}\int_{WS} \frac{\rho(\underline{r})\rho(\underline{r}')}{|\underline{r} - \underline{r}'|}\, d\underline{r}\, d\underline{r}' \qquad (5.8)$$

where there are Z valence electrons per atom. Now we assume that the pseudopotential U^{ps} can be described by

$$U^{ps}(r < R_c) = 0, \qquad U^{ps}(r > R_c) = -\frac{Ze^2}{4\pi\varepsilon_0 r}, \qquad (5.9)$$

known as the Ashcroft *empty core pseudopotential*. We have argued previously that the Coulomb attraction due to the core is largely cancelled by the "Pauli repulsion" and so this expression benefits from both mathematical simplicity and good underlying physics (if a little crudely expressed). Inserting this pseudopotential and evaluating the integrals we obtain

$$E^{tot} = Z\left\{\frac{30.1}{(r_s/a_0)^2} - \frac{12.5}{(r_s/a_0)} - (1.56 - 0.42\ln(r_s/a_0))\right\} -$$

$$\frac{40.8Z^2}{(R_{WS}/a_0)}\left[1-\left(\frac{R_c}{R_{WS}}\right)^2\right]+\frac{16.3Z^2}{(R_{WS}/a_0)} \tag{5.10}$$

again in eV per atom.

This expression has a number of features that are worthy of note. Consider sodium, for which $Z = 1$ and hence $R_{WS} = r_s$.
(i) The Hartree repulsive energy between electrons is to a large extent cancelled by the exchange and Coulomb correlation energies.
(ii) The magnitude of the electron-ion term is several eV. This is precisely the contribution we needed to make sense of the cohesive energy of metals, as discussed at the end of "jellium" section above.
(iii) The magnitude of the electron-ion attractive term falls as the core radius increases. Thus we can predict that the total energy E_{gs}^{tot} of metals will increase down a group of the Periodic Table. This is indeed the case.

Differentiating Eq. 5.10 with respect to R_{WS} to find the minimum energy, yields an expression relating the ground state value of R_{WS} and the core radius R_c. Inserting the experimentally observed WS radii we can then extract the Ashcroft core parameters R_c and also deduce the ground state energy. We find that the core radius R_c increases down a group and decreases across a period, just as we expect from rudimentary knowledge of atomic physics.

Putting in the ion cores using the WS sphere approximation allowed us to make significant progress. Our brief flirtation with the Ashcroft pseudopotential turns out to be extremely useful. We have just seen that we can explain trends in the density of metals and predict total energies semi-quantitatively. But notice that our expression for the total energy of the metal depends on R_{WS} (i.e. the density) but not the crystal structure. It's quite tricky to get this final bit of the puzzle, but we can make a simple argument to get the main point. We expect that the ions embedded into the electron gas would like to be as far away from each other as possible. For a given atomic density, this is achieved best by the BCC structure, then FCC then HCP. Thus we have explained why metals favour close-packed structures. Notice that the pseudopotential concept has provided a basis for uderstanding both the electrical properties of metals as well as their structural properties. Metals are also ductile, have low melting points, form a wide range of alloys and are soft. Can we relate these physical properties to their crystal structure?

It is not immediately obvious however that metals are crystalline at all. In 1864 Sorby showed that if you strip off the top few layers of a polished metal surface, grains typically 10^{-4} m in diameter are revealed. Each grain is a small crystal but they are randomly oriented with respect to one another - an arrangement we call *polycrystalline*. It appears that metals are not particularly fussy about forming specific bonds with specific atoms and having specific bond angles - they just want close packing. This tendency explains why metals are comparatively heavy.

As a liquid metal freezes, different seed crystals grow with a variety of orientations. The lack of specificity of metallic bonding means that when these grains meet there is no great problem. So long as the atoms at the grain boundaries can still have as many neighbours as possible, the loss of periodicity does not have a very high energy cost. This lack of specificity also accounts for the willingness of metals to form many alloys (precise chemical formulae need not

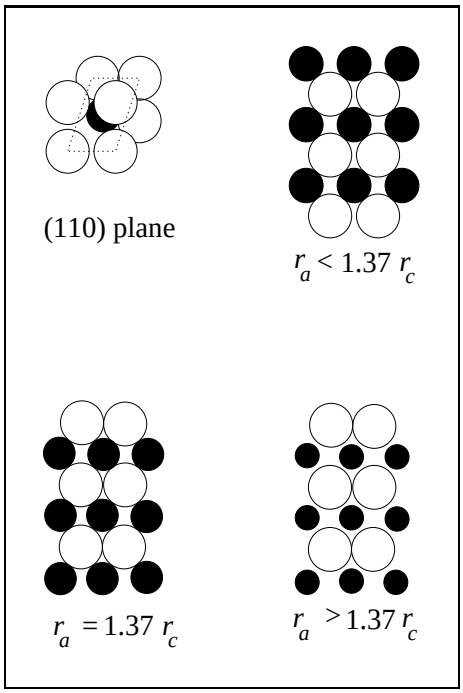

Figure 5.1: The packing of spherical ions in the (110) plane of the CsCl structure. Anions are white, cations are black.

be adhered to), their willingness to be welded together, and their low melting points.

5.1.2 Hetero-nuclear systems

For a binary AB system we distinguish three cases:

i) U_{ps}^A and U_{ps}^B are both weak.

In this case we expect the formation of metallic alloys.

ii) U_{ps}^A is strong and U_{ps}^B is weak.

In this case B will have a low ionization energy and will readily form positive ions while A will be pleased to accept an additional electron (i.e. it will have a large "electron affinity"). Ionic bonding will result. In §2.1.4 we predicted the cohesive energy (per ion pair) of a binary A^+B^- compound to be

$$E^{coh}(r) = \alpha \frac{q^2}{4\pi\varepsilon_0 r_0}. \tag{5.11}$$

r_0 is the nearest neighbour (i.e. anion-cation) distance, and $|q|$ is the net charge on each ion. The Madelung constant α is determined by the crystal structure; for the cesium chloride, sodium chloride and zincblende structures α is 1.7627, 1.7476 and 1.6381 respectively.

• *So why do all the alkali halides except CsCl, CsBr and CsI adopt the NaCl structure?*

To answer this question we must go back to our earlier observation that in the solid state the alkali halides can be considered as arrays of weakly perturbed ions. Let's assume the radii of the alkali cations r_c and halide anions r_a are constants independent of the salt.[4] Electrostatic attraction will pull these charged balls together as closely as possible, but we assume that the ions can't overlap. A (110) plane of the CsCl structure is shown in Fig. 5.1. When $r_c \approx r_a$ the oppositely charged ions can touch each other and so the anion-cation distance is $r_c + r_a$. If we now keep r_a fixed but allow r_c to shrink then it can be seen that the crystal as a whole will also shrink as the ions are allowed to get a bit closer (see Fig. 5.1). As a result the Madelung energy becomes more negative, as shown in Fig. 5.2. But if r_c is reduced beyond $r_c < 1.37 r_a$ the crystal is prevented from shrinking because the anions are now in contact as shown in the lower right panel of the figure, and so the Madelung energy cannot fall any further. When this situation prevails oppositely charged ions are prevented from getting as close to each other as they would like.[5] It can then be beneficial to adopt a more open structure which allows the oppositely charged ions to get closer together, despite the penalty of a slightly lower Madelung constant, as shown in Fig. 5.2. The critical ratios for the NaCl and zincblende structures are 2.41 and 4.45 respectively.

This line of reasoning explains why only CsCl, CsBr and CsI can adopt the CsI structure. It also explains why the nearest neighbour distances in LiCl, LiBr and LiI are not equal to $r_c + r_a$. In these salts $r_a/r_c \sim 3$ which exceeds the critical value for the NaCl structure and so the anions and cations no longer touch. The size disparity is not quite sufficient to favour the zincblende structure, however.

iii) U^A_{ps} and U^B_{ps} are both strong (but with A stronger than B).

Here, again with the exception of the noble gases, we expect covalent bonding. The mathematics should follow the discussion of the hydrogen molecule in §2.1.2 except that we must now acknowledge the distinct chemical identities of atoms A and B. For the "hetero-dimer" Eq. 2.4 for the Schrödinger equation remains valid with the potential energy term given by $U^{ps}_A(\underline{r} - \underline{R}_A) + U^{ps}_B(\underline{r} - \underline{R}_B)$ and we again seek "molecular orbital" solutions with the form $|\psi_{MO}\rangle = c_A|\psi_A\rangle + c_B|\psi_B\rangle$, where ψ_A and ψ_B are atomic orbitals corresponding to the potentials U^{ps}_A and U^{ps}_B respectively. We find eigenenergies given by

$$\epsilon^{\pm} = \bar{\epsilon} \mp \frac{1}{2}\sqrt{4h^2 + (\Delta\epsilon)^2} \tag{5.12}$$

where

$$\bar{\epsilon} = \frac{\epsilon_A + \epsilon_B}{2}, \qquad h = \langle\psi_B|\bar{U}|\psi_A\rangle, \qquad \bar{U} = \frac{U^{ps}_A + U^{ps}_B}{2}, \qquad \Delta\epsilon = \epsilon_A - \epsilon_B. \tag{5.13}$$

The corresponding wavefunctions are given by $|\psi_{MO}\rangle$ with coefficients

$$c^{\pm}_A = \frac{1}{\sqrt{2}}\sqrt{1 \pm \frac{\delta}{\sqrt{1+\delta^2}}}, \qquad\qquad c^{\pm}_B = \pm\frac{1}{\sqrt{2}}\sqrt{1 \mp \frac{\delta}{\sqrt{1+\delta^2}}} \tag{5.14}$$

[4] This turns out to be a very good approximation. Generally, cations, having lost their valence electrons, are significantly smaller than anions, which are slightly bloated due to their excess electrons.

[5] There are a few cases where the cations are larger than the anions. In this case we just reverse the roles of r_c and r_a.

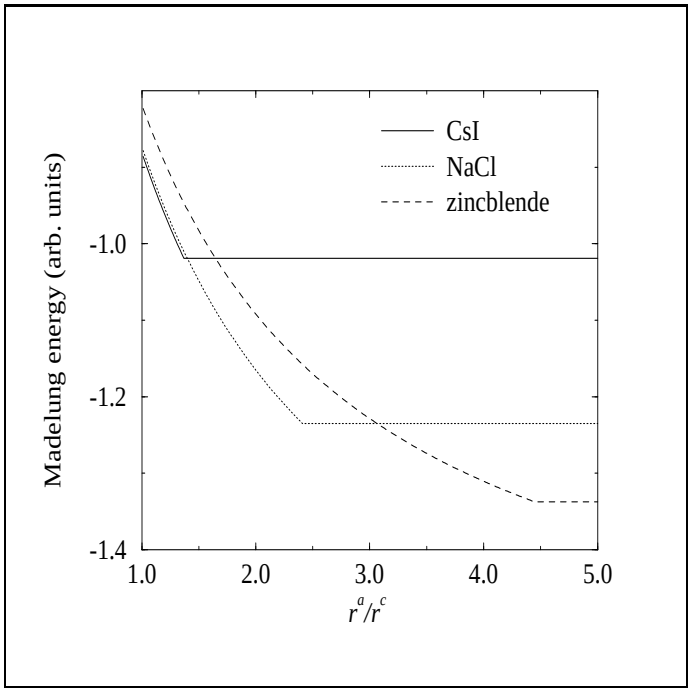

Figure 5.2: The Madelung energy for the CsI, NaCl and zincblende structures as functions of r_a/r_c, where r_a is held fixed. It can be seen that as the ratio increases it can become more favourable to adopt structures with smaller Madelung constants.

where $\delta = \Delta\epsilon/(2|h|)$. As seen for the homonuclear case considered in §2.1.2 we see here the formation of a bonding state with lower energy than $\bar{\epsilon}$ and an antibonding state with higher energy. Notice that now we have an ionic contribution to the bonding-anti-bonding splitting (Eq. 5.12). Also the bonding state has greater electron density on atom A (with the stronger pseudopotential) whereas the anti-bonding state has more density on the less favoured B site. Thus we have mixed ionic-covalent bonding.

It is interesting to compare the isolelectronic solids Ge (Group IV), GaAs (III-V), ZnSe (II-VI), and CuBr (I-VII). Ge is totally covalent and adopts the diamond crystal structure, each atom forming four covalent bonds with its nearest neighbours. The III-V and II-VI solids also tend to exhibit fourfold coordination but are slightly ionic, and this is reflected by the increased band gaps (shaded pink regions) in Fig. 4.19. For the I-VII compounds the bonding is predominantly ionic in character and to a good approximation all the valence electrons reside on the halide ions. The covalent bonding charge is so depleted that it can no longer hold the four-fold coordinated structures together. As we have already seen, ionic salts adopt structures which maximise their Madelung energy. It is observed that $\delta = 1.24$ separates all the binary compounds into four-fold coordinated ($\delta < 1.24$) and ionic structures ($\delta > 1.24$).

The bonding in most ceramics and minerals has a significant covalent component. Although covalent solids have well-defined chemical formulae, a surprising range of structures are observed for many compounds, and we can explain this. We have already said many times that Si likes to bond to four other atoms, forming little tetrahedra. Tetrahedra can link together in a variety of ways. E.g. SiO2 subunits can link to form a fibrous chain (as in asbestos), or laminar structures (as in talc or mica), or as 3D structures (as in quartz). Each has a well defined chemical structure, but the flexibility of the tetrahedral unit explains the diversity of forms silicon dioxide can take.

The nature of covalent bonding dictates the formation of molecules and solids with specific chemical compositions and bond angles. The strength of these bonds ($\sim$ 1 eV per bond) explains the high melting points of many ceramics, and their directionality means that any attempt to distort such a crystal will meet with extreme resistance, i.e. covalent solids are very rigid. Planes of atoms are unable to slip past each other since the energy barrier for sheer distortion is extremely high. If a sufficiently strong force is applied to cause significant distortion then the chemical bonds are broken and the solid fractures i.e. covalent solids are brittle.

5.2 A "first principles" approach to cohesion

This section is not really difficult but requires us to think carefully about what we have been doing. To start with look again at Eq. 4.1, the Schrödinger equation for the electronic system of a solid. This gives an eigenvalue equation for Ψ, the wavefunction of the electronic system, with electronic energy eigenvalue, E. But this is not the whole energy of the solid since it does not include the potential energy due to the ion-ion repulsion or the ionic kinetic energy. Furthermore we have assumed the set of atomic positions $\{\underline{R}\}$ is known, and that they don't vary. Let's go back a step and try to clarify the relation between the electronic and ionic degrees of freedom.

5.2.1 The Schrödinger equation and Born-Oppenheimer separation

The Schrödinger equation describing both the electrons and the nuclei of a solid is

$$\mathcal{H}\Phi(\{\underline{r}\},\{\underline{R}\}) = \left\{\sum_j -\frac{\hbar^2}{2M}\nabla_j^2 + \frac{1}{2}\sum_{k,j(k\neq j)}\frac{Z^2e^2}{4\pi\varepsilon_0|\underline{R}_j-\underline{R}_k|} + H\right\}\Phi(\{\underline{r}\},\{\underline{R}\})$$

$$= E^{tot}\Phi(\{\underline{r}\},\{\underline{R}\}) \tag{5.15}$$

where Φ is the wavefunction of the entire system, M and Ze are the mass and charge on the ions[6], $\{\underline{r}\}$ is the set of electron coordinates, $\{\underline{R}\}$ is the set of ion coordinates.[7] The first two terms in the second equality are the ionic kinetic energy and the inter-ion potential energy, and H is the electronic Hamiltonian (see Eq. 4.1) for a particular set of ion positions:

$$H = \sum_i\left\{-\frac{\hbar^2}{2m}\nabla_i^2 - \sum_j\frac{Ze^2}{4\pi\varepsilon_0\,|\underline{r}_i-\underline{R}_j|}\right\} + \frac{1}{2}\sum_{i,j(i\neq j)}\frac{e^2}{4\pi\varepsilon_0\,|\underline{r}_i-\underline{r}_j|}. \tag{5.16}$$

Notice that in general the wavefunction of a solid depends on the coordinates of all its electrons and all its ions. If both electrons and ions are moving, the concept of electronic structure is strictly speaking not well-defined. In practice, electronic and ionic motions are separable to a good approximation since the mass of the electron m is extremely small compared to ionic mass M (m/M is between 10^{-4} and 10^{-5} in most cases). The Born-Oppenheimer approximation claims that since electrons can respond much faster than ion cores, then even if the ions are moving the electrons *instantaneously* adjust. In other words, the electronic system is in the ground state appropriate for the particular ion configuration at any particular instant. In quantum mechanics the smooth variation of an external potential is known as an *adiabatic perturbation* and leads to a smooth evolution of the eigenstates of a system and not a mixing of the eigenstates.

Born-Oppenheimer separation, often called the *adiabatic approximation*, suggests that we look for eigenstates of the full Schrödinger equation (Eq. 5.15) which have the form

$$\Phi(\{\underline{r}\},\{\underline{R}\}) = \Psi(\{\underline{r}\},\{\underline{R}\})\Upsilon(\{\underline{R}\}) \tag{5.17}$$

where Ψ is the wavefunction of the electronic system for a given ionic configuration, and Υ is the wavefunction of the ions. Inserting this expression into Eq. 5.15 we obtain

$$\left\{\sum_j -\frac{\hbar^2}{2M}\nabla_j^2 + \frac{1}{2}\sum_{k,j(k\neq j)}\frac{Z^2e^2}{4\pi\varepsilon_0|\underline{R}_j-\underline{R}_k|} + E(\{\underline{R}\})\right\}\Upsilon(\{\underline{R}\})\Psi(\{\underline{r}\},\{\underline{R}\}) =$$

[6] As previously we are lumping the nucleus core electrons of an atom together to form an ion. For notational simplicity we consider here solids which contain only one type of atom.

[7] These are not necessarily Bravais lattice points since the ions can move.

$$\Psi(\{\underline{r}\},\{\underline{R}\})\left\{\sum_j -\frac{\hbar^2}{2M}\nabla_j^2 + \frac{1}{2}\sum_{k,j(k\neq j)}\frac{Z^2e^2}{4\pi\varepsilon_0|\underline{R}_j-\underline{R}_k|} + E(\{\underline{R}\})\right\}\Upsilon(\{\underline{R}\})+$$

$$\left\{\sum_j -\frac{\hbar^2}{M}\nabla_j\Upsilon.\nabla_j\Psi + \Upsilon\sum_j -\frac{\hbar^2}{2M}\nabla_j^2\Psi\right\} = E^{tot}\Upsilon(\{\underline{R}\})\Psi(\{\underline{r}\},\{\underline{R}\}) \quad (5.18)$$

where $E(\{\underline{R}\})$ is determined by

$$\left[\sum_i\left\{-\frac{\hbar^2}{2m}\nabla_i^2 - \sum_j\frac{Ze^2}{4\pi\varepsilon_0\,|\underline{r}_i-\underline{R}_j|}\right\} + \frac{1}{2}\sum_{i,j(i\neq j)}\frac{e^2}{4\pi\varepsilon_0\,|\underline{r}_i-\underline{r}_j|}\right]\Psi(\{\underline{r}\},\{\underline{R}\})$$

$$= E(\{\underline{R}\})\Psi(\{\underline{r}\},\{\underline{R}\}). \quad (5.19)$$

Now consider the final two terms on the left hand side of Eq. 5.18. If we ignore them then the ionic wavefunction satisfies

$$\left\{\sum_j -\frac{\hbar^2}{2M}\nabla_j^2 + \frac{1}{2}\sum_{k,j(k\neq j)}\frac{Z^2e^2}{4\pi\varepsilon_0|\underline{R}_j-\underline{R}_k|} + E(\{\underline{R}\})\right\}\Upsilon(\{\underline{R}\}) = E^{tot}\Upsilon(\{\underline{R}\}) \quad (5.20)$$

While the two are clearly inter-related, we have achieved a separation of the ionic and electronic motion. To be explicit:

The ionic coordinates are parameters in the electronic Schrödinger equation.

The energy from the electronic Schrödinger equation acts as a potential term in the Schrödinger equation describing the ionic motion.

$E(\{\underline{R}\})$ is the adiabatic contribution of the electrons to the total energy. It describes the electronic ground state appropriate to the prevailing ionic configuration denoted by the set $\{\underline{R}\}$, and is independent of the ionic velocities. Returning now to the neglected non-adiabatic terms we note that in FOPT these terms do not affect the total energy and we are justified in omitting them. However, they do affect the wavefunction, or equivalently they cause transitions between the adiabatic eigenstates. This is the electron-phonon scattering process which we spoke of in the previous chapter.

5.2.2 The potential energy hypersurface

The two potential energy terms in Eq. 5.18 give the *Born-Oppenheimer potential energy hypersurface* V^{BO}:

$$V^{BO} = \frac{1}{2}\sum_{k\neq j}\frac{Z_iZ_je^2}{4\pi\varepsilon_0|\underline{R}_j-\underline{R}_k|} + E(\{\underline{R}\}). \quad (5.21)$$

Here we are primarily interested in determining the cohesive energy and so we take $T = 0$. Neglecting the zero point motion of the ions, we can assume $E^{tot}(gs)$ is given by the minimum of the BO potential, so to find the ground state energy of a solid we must find the ionic coordinates corresponding to the energy minimum of the Born-Oppenheimer potential energy hypersurface.

5.2.3 The ground state energy

Since the ion-ion term in Eq. 5.21 is straightforward, it remains only to recall an expression for the electronic energy E. Unfortunately we never actually derived such an expression since the Schrödinger equation for the whole electronic system is a many-body problem and therefore insoluble. We made progress by making the one-electron approximation, which meant solving simpler equations for ψ, the wavefunctions of single electrons, and ϵ, the corresponding eigenenergies. Now we need to specify how the one-electron eigenenergies relate to the total electronic energy. Since the electrons interact, this is not trivial. The equations we need are

$$E = T_{el.} + V_{el.-ion} + V_{el.-el.}, \qquad E_{bs} = \sum_j \epsilon_j = T_{el.} + V_{el.-ion} + 2V_{el.-el.}. \tag{5.22}$$

If we sum the one-electron eigenvalues to obtain the "band-structure energy" E_{bs} then we count each electron-electron interaction twice.[8] It follows that

$$E_{gs}^{tot} = E_{bs} + V_{ion-ion} - V_{el.-el.}. \tag{5.23}$$

$V_{el.-el.}$ is already part of the band structure calculation and $V_{ion-ion}$ is relatively easy to deal with and so we can readily adapt our band structure flow diagram from §1 so as to obtain E_{gs}^{tot} by varying the set $\{\underline{R}\}$ (i.e. we vary of the crystal structure and nearest-neighbour distance). An extended flow diagram is shown in Fig. 5.3 and results of such a calculation for Si are shown in Fig. 5.4. A lot of information can be deduced from this figure:
(i) The diamond crystal structure is predicted.
(ii) The energy minimum occurs at an atomic volume in extremely good agreement with experiment (i.e. we can predict the bond length).
(iii) Subtracting E_{∞}^{tot} from the ground state energy of the solid (see Eq. 5.1) gives a cohesive energy of 4.67 eV per atom, in excellent agreement with the experimental value of 4.63 eV/atom.
(iv) The curvature of the energy curve near the ground state (i.e. the second derivative with respect to volume) is essentially the stiffness of the bonds which in turn determines vibrational frequencies and the bulk modulus. These quantities are also also turn out to be in excellent agreement with experiment.
(v) Under compression (i.e. on reduction of the atomic volume by application of an external force) the diamond structure is not necessarily favoured. Thus we have a method for calculating the *phase diagrams* of solids from first principles. The pressure-induced transition to the β-tin structure has been confirmed experimentally for silicon.

The original motivation for the development of band theory in the 1930's was the desire to explain the cohesion of solids. Although band theory has long been able to describe non-equilibrium phenomena (e.g. optical absorption, electrical conduction etc.), the calculational scheme described in this section has only yielded quantitative predictions of the cohesive energies of solids in the last decade or so. We can see two major reasons for this. Firstly, the success

[8] The one-electron Schrödinger equation for electron i contains the interaction from electron j, and the one-electron Schrödinger equation for electron j contains the interaction from electron i. The many-electron Schrödinger equation contains the interaction between electron i and electron j only once.

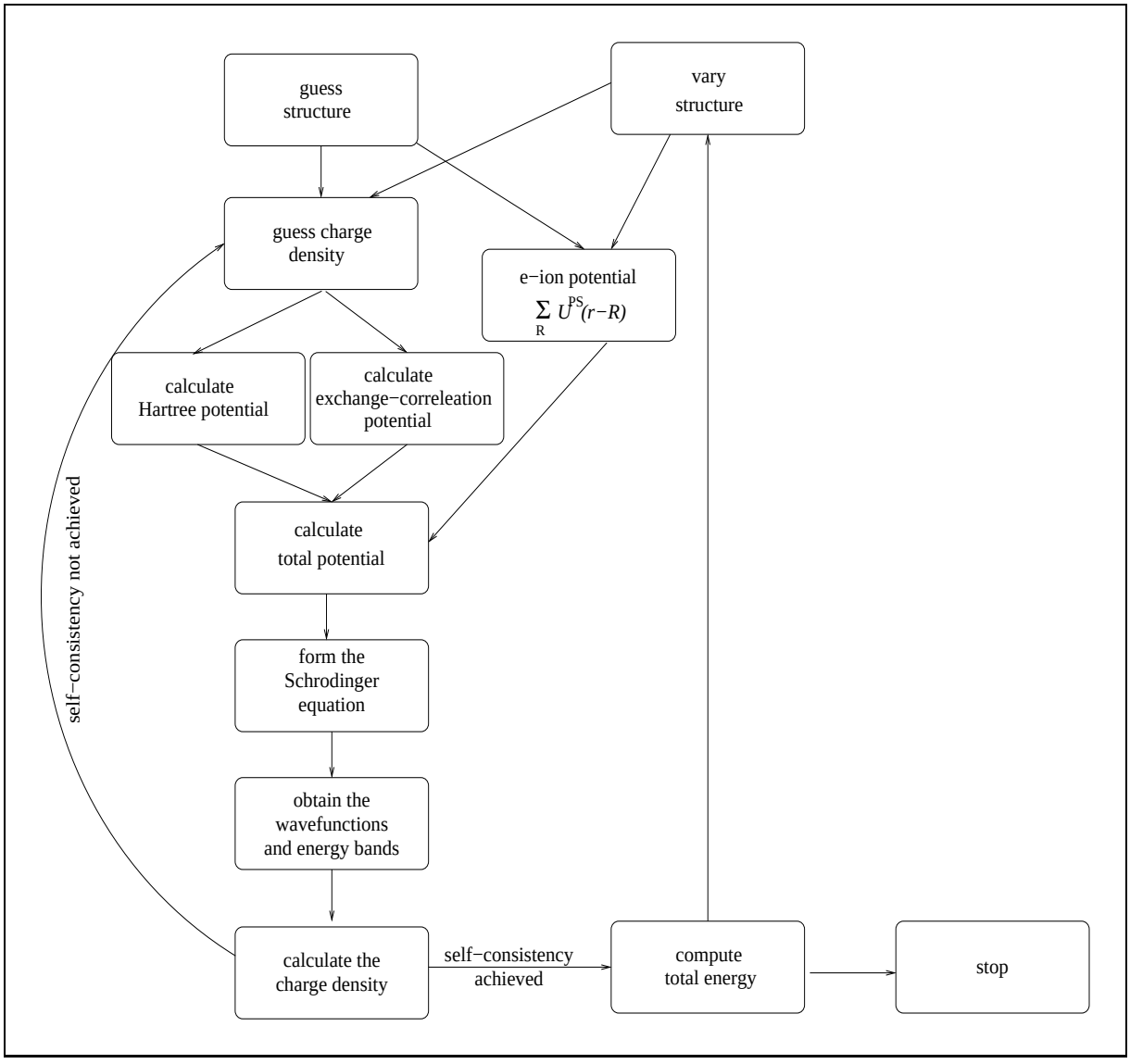

Figure 5.3: Calculational scheme for obtaining E_{gs}^{tot} of a solid. The electron-ion interaction is now written explicitly.

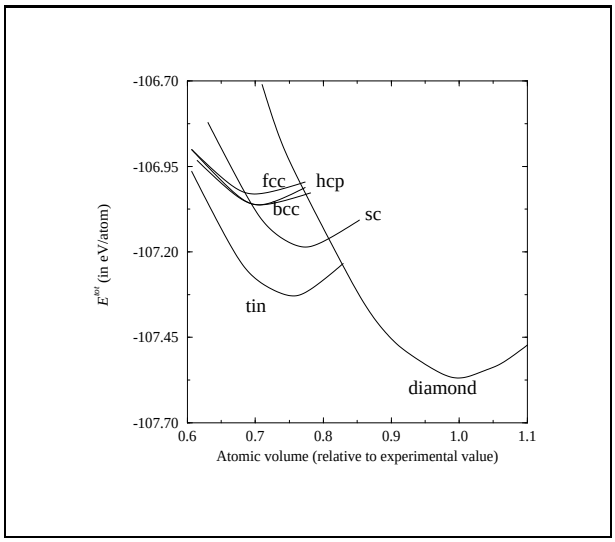

Figure 5.4: Total energy calculations for Si in a variety of crystal structures as functions of density.

of density functional theory in representing electron-electron interactions was only apparent at around 1980. Secondly, supercomputers are needed in order to perform the enormous number of self-consistent calculations represented in Fig. 5.4.

5.2.4 Atomic dynamics

It is well worth repeating the point from §5.2.1 that the electronic energy E, together with the inter-ion repulsion, creates the potential which controls the atomic dynamics in a solid. The flow diagram in 5.3 is therefore a single theoretical framework describing electronic structure, crystal structure and crystal dynamics, the three main branches of solid state and condensed matter physics.

- *We now have a truly first principles quantum theory of the solid state.*

This subsection is a digression, but we can include a few other bits and pieces here for completeness.

The Hellmann-Feynman theorem

Consider first a ball which can move in the xy plane. Suppose we calculate $V(\underline{r})$ the potential energy of the ball at each point in the plane and find the results illustrated by the contour map in Fig. 5.5. Since we know the energy at each point we can deduce the force the ball would experience at each point using the expression

$$\underline{F} = -\frac{dV}{d\underline{r}}. \tag{5.24}$$

This means that the force on the ball at some point tends to accelerate it in the direction for which the energy decreases fastest at that point. Knowing $V(\underline{r})$,

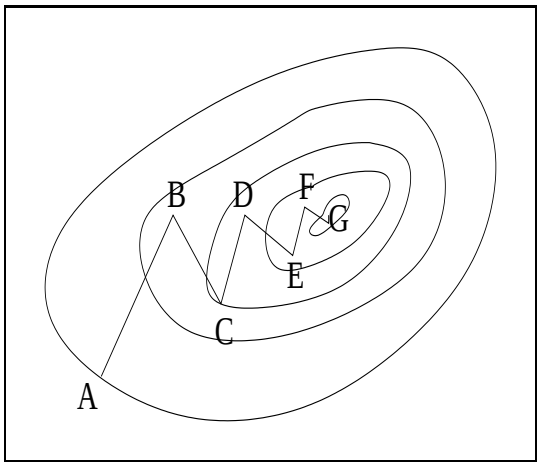

Figure 5.5: Potential energy contours for a ball moving in the xy plane. The potential energy decreases with the radius of the contours. The path shown is a zero temperature quench, not a physical trajectory.

it should then be possible, given an initial position and velocity, to calculate trajectories of the ball moving across the plane. Suppose the ball starts at position $\underline{r}_0$ at time t_0 with velocity $\underline{v}_0$. A small time Δt later, the position and velocity of the ball will be

$$\underline{r}_1 = \underline{r}_0 - \frac{(\Delta t)^2}{2M}\left(\frac{dV}{d\underline{r}}\right)_{\underline{r}=\underline{r}_0}, \quad \underline{v}_1 = \underline{v}_0 - \frac{\Delta t}{M}\left(\frac{dV}{d\underline{r}}\right)_{\underline{r}=\underline{r}_0}. \tag{5.25}$$

We can then calculate the force on the ball at the new position and so predict its position after a second time step. In this way we can plot the trajectory of the ball, provided small time steps are used.[9]

In fact we can be much more economical than this. Rather than calculating V at all positions in advance, let's just start by calculating it in the vicinity of $\underline{r}_0$. This is all the information we need to determine the trajectory during the first time step. Having determined $\underline{r}_1$ we then calculate V around this point, and so on. Thus, to calculate the trajectory of the ball we only need to do energy calculations along the trajectory itself and very close to it, not everywhere.

We can apply the same procedure to calculate atomic motions, assuming they satisfy Newtonian dynamics. The force on atom i is

$$\underline{F}_i = -\frac{dV}{d\underline{R}_i} = -\frac{\partial V}{\partial \underline{R}_i} - \frac{\partial V}{\partial \Psi}\frac{d\Psi}{d\underline{R}_i}. \tag{5.26}$$

Notice that the use of the total derivative in the first equality suggests that we must allow for the relaxation of the electrons produced by changes in atomic coordinates, as explicitly shown in the second equality. According to the Hellmann-Feynman theorem this is not necessary after all, because we can start by calculating the force at some point the Born-Oppenheimer surface, i.e. when the

[9] Eq. 5.25 assumes that the acceleration is constant during each time step. This approximation deteriorates as Δt increases.

electronic system is in its ground state for a given set of nuclear coordinates (which means $\partial\Psi/\partial\underline{R}_i = 0$). Under these circumstances the second term in the second equality above must be zero, and we obtain

$$\underline{F}_i = -\frac{\partial V}{\partial \underline{R}_i}. \tag{5.27}$$

This means that to calculate the force at some point on the Born-Oppenheimer surface we need only determine how the energy changes with respect to variation of the atomic coordinates at that point without the need to keep relaxing the electrons. As soon as we move the atoms a finite distance according to their Hellmann-Feynman forces we in evitably move away from the Born-Oppenheimer surface and so we lose accuracy. It is therefore necessary to regularly update the electronic wavefunctions, but not at every timestep.

The dynamics of diffusion, chemical reactions, lattice vibrations etc. can all be studied from first principles in this way, but it is extremely computationally demanding.

Spring constants and empirical potentials

In §3 we explored lattice dynamics with rather crude ball and spring models. We now have the means to determine the spring constants needed for these models. Of course we can do better by using the real Born-Oppenheimer potential energy surface in place of the model. A particularly useful approach is to use the full computational machinery in Fig. 5.3 to deduce empirical interatomic potentials. These can then be used to predict atomic dynamics in systems too complex for full electronic structure calculations to be viable. In this way one can hope to combine the speed of a model calculation with the accuracy of the full machinery (since the quantum mechanics is in some sense built into the empirical potential).

Finding the ground state

We saw above that the calculation of atomic dynamics from first principles is made computationally tractable by the Hellmann-Feynman theorem and by only performing calculations along the trajectory itself. If we are only interested in finding the ground state we can make much greater economies.

Let's consider the ball again. If we calculate the energy at every point we can decide what is the lowest energy position. In Fig. 5.5 the ground state is the point G. This is analogous to predicting the ground state structure of Si from the huge amount of computational effort required to produce Fig. 5.4. Now let's use the dynamics approach to make a short cut. Let's start with the ball stationary at A and calculate the force at that point. Now let's calculate the energy at a series of points along the direction of that force until the energy starts to increase. We stop at this point (B) and recalculate the force. We then start calculating the energy along the direction of this new force, rapidly converging towards the ground state G. This approach allows us to move to the ground state with very few energy calculations.

Note that the route to the ground state is not a physical trajectory of the ball for two reasons.
(i) the acceleration on the ball was only calculated correctly at the points A, B,

etc. Along the line AB there is no guarantee that the force on the ball is even along this direction.
(ii) the procedure described amounts to picking the ball from point A and placing it at rest at point B then C then D etc. In other words the ball acquires no kinetic energy on this strange journey. This fictitious trajectory is called a "zero temperature quench".
Provided we are interested only in locating the ground state we can treat the description of the ball's trajectory as sloppily as we wish in order to speed up the calculation. By calculating the Hellmann-Feynman forces at a few points on the Born-Oppenheimer surface, zero-T quench dynamics can also be applied to find the ground state atomic positions of the atoms in a solid.

This is a brief guide to the main concepts in each chapter of the course. You may wish to use this summary to structure your revision since the assessment of the course will be aligned with the main concepts. But note that small parts of the exam paper may draw upon material designated as primarily "background" and even previously unseen material.
You may find it useful to get exam practice by looking at the papers from 2001-2008, however do note that the syllabus has evolved a little over the years. The "mock exam" (problem sheet 6) corresponds to the current syllabus.

Chapter 1.

Drude model (classical free electron gas)
The physical picture of the collision model and the assumptions on which it is based.
MB velocity distribution and estimate of core radius ($\sim$ 1 Å) lead to $\tau \sim 10^{-14}$ sec. at room temperature.
Derivation of Ohm's law from Newton's second law (Eq. 1.4).
Electronic contribution to the heat capacity (Eq. 1.10) comes out wrong.

Sommerfeld model (quantum free electron gas)
Stationary states of SE are plane waves with wavevector $\underline{k}$.
Periodic boundary conditions give allowed $\underline{k}$ and hence $g^{\underline{k}}$.
How to convert $g^{\underline{k}}$ to g^k and g^ϵ.
Ground state of free electron gas obtained by populating the allowed states in ascending order of energy, the maximum energy being the Fermi energy.
Characteristic properties of the Fermi gas: $\epsilon_f \sim$ a few eV, $k_f \sim 1$ Å^{-1}, $v_f \sim 10^6$ ms^{-1}, $T_f \sim 10^5$ K.
Occupancy at $T > 0$ given by FD function.
The FD and MB distributions are spectacularly different at room temperature, and this has a radical effect on the electrons' contribution to the heat capacity. We derived a good approximation to this in Problem Sheet 1, but there is a short cut - the result for a classical gas ($\frac{3}{2}nk_b$) is reduced by a factor $\sim k_bT/\epsilon_f$ since only this fraction of the electrons can be thermally excited (page 13). (Remember that the electrons' contribution to C_V can only be seen at low temperature. At higher T it is swamped by another effect.)
Sommerfeld model still needs the collision idea to explain finite conductivity, but, since using $\tau \sim 10^{-14}$ s in Eq. 1.5 gives agreement with experiment, we must conclude the collision length is $\sim v_f\tau$ = hundreds of atomic spacings at room temperature. So what is the true origin of collisions?

Chapter 2.

Structure
§2.1-3 are included mainly to provide background and context.
The description of a crystal structure in terms of a BL and a basis, together with related ideas (unit cell, packing fraction etc.) are vital (see §2.4). Familiarity with the specific examples discussed in §2.4.3 is expected.

Reciprocal lattice
This is a key piece of machinery, essential to much of the course, so know and understand Eq. 2.20-25 thoroughly.
The meaning of the term "Brillouin zone" is important.
The importance of the reciprocal lattice lies in the fact that a plane wave with wavevector equal to a reciprocal lattice vector will have the periodicity of the real space lattice (i.e. will have the same value at every lattice point) as shown in Fig. 2.25. You should know the theorem

connecting planes and reciprocal lattice vectors on page 51.

Diffraction
A theoretical description of the diffraction of x-rays from a crystal must address: (i) the condition for diffraction to occur, and (ii) the amplitude (and hence intensity) of the diffracted beams. §2.6 deals with the former, §2.7 the latter.
You should know that the condition for diffraction to occur can be expressed using the Bragg equation (Eq. 2.28, Fig. 2.26), the von Laue equation (Eq. 2.29 and Fig. 2.27), or in terms of Bragg planes (Eq. 2.30 and Fig. 2.28), and you should appreciate that these are equivalent.
The key equations in §2.7 are 2.39-40 (don't worry about factors of V in these equations). (Remember that experiments measure the it intensities of diffraction spots while structure factors determine their *amplitudes*.)
You should know and understand the variation of the atomic form factor with Q (and number of electrons - we usually assume atoms in solids are spherical), and be able to find expressions for the structure factor of a given crystal structure.
§2.8 is background material.

Chapter 3.

Classical vibrations
What is meant by the "harmonic approximation".
The equations of motion of a set of coupled harmonic oscillators can be rearranged to give a set of uncoupled equations, each describing a (possibly fictitious) harmonic oscillator. These "normal modes" are independent.
Be able to derive the dispersion relations for the monatomic and diatomic chains (Eq. 3.13 and 3.17).
Know Fig. 3.4 in detail and be able to explain its features.
Similarly, know Fig. 3.7.
Periodic boundary conditions and counting the modes (also know how many modes in 3D).
The "branch structure" of materials with different chemical bonding (but the same structure) is quite similar, except the frequency axis is rescaled, as in Fig. 3.9.

Quantised vibrations and phonons
The general discussion of the quantum theory of the harmonic crystal in §3.4 is important.
Know and understand the shape of C_V as a function of T for a solid.
Understand the Einstein and Debye models of the vibrational density of states.
§3.6-8 are mainly background discussion, but you should be aware of the conservation rules for inelastic scattering from a vibrating crystal (Eq. 3.40-41).

Chapter 4.

In general
Appreciate that the idea of solving the SE "one electron at a time" is an approximation (expressed in Eq. 4.2).
Be familiar with how dispersion curves (i.e. ϵ vs $\underline{k}$ for electrons, ω vs $\underline{k}$ for phonons) are plotted for crystals.
Know how many electrons each energy "band" can accommodate (using periodic boundary conditions).
Filling of bands is non-trivial for 2D and 3D (know and understand Fig. 4.3 thoroughly).

NFEM

The primary effect of a weak crystal potential U is to cause diffraction when the electron wavevector lies on a Bragg plane. The result is "band gaps" (§4.3.1). Appreciate the important consequence of this fact, explored in §4.3.2-3.

Realistic band structure
know Bloch's theorem (Eq. 4.11), the properties of Bloch electrons (§4.4.2), and the justification for use of the Reduced Zone Scheme for electrons (proof on p107).
Know that the electronic SE can be solved "in k-space" (§4.5) but that the core orthogonal constraint presents a technical challenge (§4.6.1).
Appreciate the value of the U^{ps} idea in resolving this problem and in explaining chemical trends, existence of free electron metals etc. (§4.6.3 and part of §5.1).
§4.6.2 and §4.6.4 are background only.
Understand the Hartree "self-consistent field" approach to treating the (many body) electron-electron interaction. §4.7.2-5 are mainly background reading, but you should understand that exchange and Coulomb correlation are important to the stability of metals (see also 5.1.1).

Electron dynamics between collisions
Know and understand §4.8.
Filled bands are inert.
The concept of "holes" is important (§4.9).
Interband optical transitions are vertical on account of the conservation rule in Eq. 4.54.

Origin of collisions
In band theory the electron-ion interaction (in the perfect crystal) is fully accounted for when solving the Schrodinger equation (i.e. the Drude mechanism is wrong, as we suspected).
Ions whose position departs from that in the ideal crystal structure do scatter electrons. This circumstance includes structural defects, chemical impurities, and also lattice vibrations. Electron-phonon scattering dominates in pure crystalline specimens, provided there is a significant number of phonons (i.e. for T approaching the Debye temperature).
(Technically electon-electron interactions must also be considered, but this turns out to be unimportant.)

Semiconductors
Concentrate on §4.12.1.
Understand the basic mechanism for electrical conduction in semiconductors.
Different "Bloch" states have different effective masses and hence different mobilities.
Since carrier densities are low FD statistics approach the classical (MB) limit.

Insulators and superconductors
§4.13 − 4.14 are background reading only.

Chapter 5.

Cohesion
§5.1: this is mostly background discussion, except that you should appreciate the ingredients needed to explain the stability of metals and understand the discussion of competing ionic structures.
§5.2: know the basic argument underpinning Born-Oppenheimer separation.

Condensed matter physics

Dr. R.J. Cole
Department of Physics and Astronomy, University of Edinburgh

October 9, 2002

Chapter 0

Introduction

0.1 What is Condensed Matter Physics?

Matter is the physical expression of the laws of nature. In this course we aim to explain the properties of 'condensed' matter, by which we mean matter with a density of $\sim 10^{-3}$ kg m^{-3}. Given the enormous diversity of materials, this is a tall order indeed, and so most of the time we will have to be content with a qualitative or semi-quantitative approach for just a few properties. I hope that you will be convinced that condensed matter physics is intellectually demanding, reasonably interesting, and is still an extremely active research area, particularly at Edinburgh.

There are three main components to the course:
- Atomic structure (i.e. the spatial arrangement of atoms),
- Lattice dynamics (i.e. how the atoms move), and
- Electronic structure (i.e. how the electrons behave in condensed matter).

Rather than present a systematic exposition of each, here we adopt an idiosyncratic approach. We will gain insight into mechanical, thermal and optical properties but will focus on the material property that shows the greatest variation, namely electrical conductivity, σ: that of the best conductors is about 10^{30} times that of the worst. If we understand anything about condensed matter then we must be able to explain this.[1]

So profound is the variation in σ that one might suspect that completely different conduction mechanisms are involved in different materials, and it is so. The best electrical conductors, which we call metals, offer the simplest starting point. We may suppose that for an electrical current to flow through a metal there must be a collection of mobile electrons within these materials, and we will pursue this notion in the next chapter. The casual explanation for the existence of insulators is that their electrons are tied up in rigid chemical bonds. We will see that this is a rather superficial (and in fact misleading) explanation.

[1] It is worth noting at the outset that many of the properties of condensed matter appear to be linked. For example those solids which crystallise with what we call the diamond crystal structure tend to be physically hard, electrically insulating with high melting points. Elements which exhibit close-packed structures are usually good conductors of heat and electricity, shiny, ductile, and rather soft.

0.2 Some preliminary considerations

0.2.1 A very many body problem

Whether by classical or quantum mechanics, we are very good at solving problems where a single particle is moving in an external potential (e.g. an electron moving in an electric field) or when two particles interact with each other (e.g. the earth going round the sun). But we cannot in general *solve* the equations of motion for a system of only three interacting particles. Condensed matter comprises lots of particles, all of which are interacting with each other. Although we understand these interactions (and of the fundamental forces we shall have need only of the electromagnetic interaction) this "many-body" nature makes theoretical treatments hard. How then are we to approach condensed matter systems? A good physicist must master the art of carefully applying approximations.

One approach is to focus on the movement of one particular particle. We can then assume that all the other particles give rise to some average potential which acts upon our chosen particle. We can then compute the movement of this particle in the presence of this potential (which we must guess since we don't know how all the other particles move of course). In this way we are approximating the behaviour of the many-body system by that of a collection of single particles which, though interacting, move independently. It turns out that this simplification enables us to understand and explain many things. But occasionally we find phenomena that can only be explained by looking at some of the correlation effects inside the many-body system. Superconductivity is one such example. Who would have guessed that an apparent lack of electrical resistance in a metal could be produced by electrons attracting each other? We all know that electrons repel each other, but in a superconductor they appear to bound together in pairs.

Such effects only emerge from the many-body system. We cannot predict them by looking at how two particles interact or how a single particle moves in an external field. This disturbs the reductionist approach of trying to understand the universe by dissecting it.

0.2.2 Quantum or classical, particle or wave?

Quantum mechanics can be tough, so can we get away without it? We know that when a beam of light is fired at a narrow slit then diffraction effects can be observed - the wave nature of light is manifest. In fact the de Broglie relation tells us that any particle with momentum p can be thought of as a wave with wavelength $\lambda \sim h/p$. But when we walk through a doorway we tend not to notice any diffraction. So how can one decide if a particle will behave classically,[2] or if its wave nature will dominate (requiring us to solve the Schrödinger wave equation)?

The important point is whether the de Broglie wavelength is comparable in magnitude with the relevant physical size for the situation in question (which might be the width of a slit, the separation between two particles etc.). If the wavelength is negligible (e.g. when you walk through a doorway at 1 m s^{-1})

[2] By "classical" we mean "non-quantum mechanical" here. We will probably not encounter any relativistic effects.

then classical physics works, if not, then we must deal with waves. You might be a bit worried that if we sufficiently reduce the momentum of an object as it approaches an aperture, then it would always be possible to make the de Broglie wavelength significant. However it is not always be possible to make the de Broglie wavelength arbitrarily large for a variety of reasons, and here we can list three.
Firstly, there may be no suitable allowed states.
Secondly, an object with temperature T will have a random motion with energy of order $\sim k_b T$, where $k_b = 1.38 \times 10^{-23}$ J K^{-1} and is known as Boltzmann's constant, and hence a "thermal wavelength" of $\sim h/\sqrt{2mk_bT}$. The thermal wavelength of a snooker ball at room temperature is $\sim 10^{-23}$ m, which is extremely small on the length scale of the ball, and we confirm that wave effects (i.e. quantum mechanics) should not be manifest during a game of snooker, no matter how slowly you strike the balls. On the atomic scale, where here the natural length is 1 Å $= 10^{-10}$ m, the same argument implies that at room temperature the molecules in water are bumbling around like classical particles, while the electrons within the molecules demand a quantum mechanical treatment. It may seem that reducing T will eventually cause the de Broglie wavelength to explode, but there is a third point.
The Uncertainty Principle requires that the de Broglie wavelength cannot exceed Δx, the spatial uncertainty. This implies that atoms in condensed matter can usually be treated classically.

0.2.3 Statistics

Quantum mechanics is not simply a matter of finding the eigenstates of a Hamiltonian. When we consider a system of many particles we must think about statistics. Take, for example, the electrons in a piece of metal. If, using some approximation, we describe this many body system using a collection of single electron states, it still remains for us to say how many electrons will go into which states. At $T = 0$ they will try to drop as low in energy as they can, and we call this the ground state. For $T > 0$ there will inevitably be thermal excitations, and we also need to understand how this affects the occupations. In classical systems the probability that a particular state will be occupied is given by Maxwell-Boltzmann statistics but in quantum mechanical systems things are more complicated. It is a remarkable fact that all systems of particles with half-integer spin obey one kind of statistics (Fermi-Dirac statistics), while those consisting of integer spin obey another (Bose-Einstein statistics). This "spin-statistics theorem" is universal - no exceptions are known.

Already we have touched upon some rather profound issues and it should be clear that if we are to get to grips with condensed matter then we will need to keep our wits about us. Since condensed matter constitutes the raw materials for all manner of industries from food production to microelectronics to ship building, the technological imperative for mastering this field can scarcely be overstated.