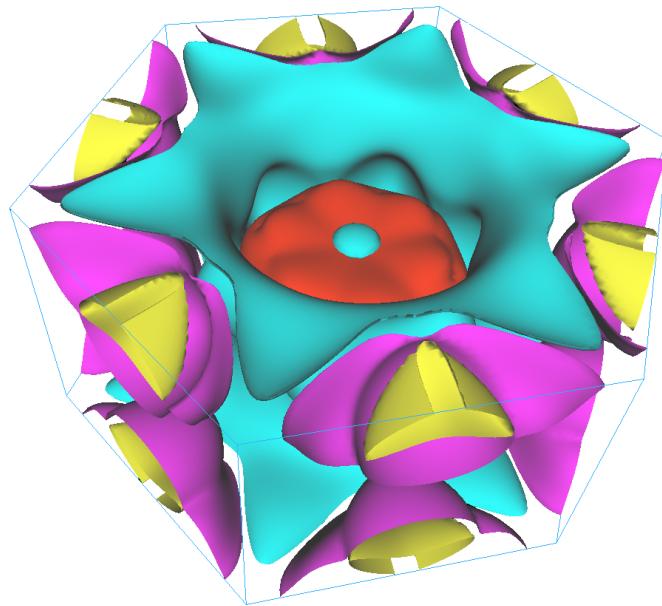# PHYS3004
Crystalline Solids

Cover image shows the Fermi surface of Technetium (Tc)

Preface

Condensed matter physics is concerned with the study of matter in a solid form as found in the world around us. It is often considered synonymous with the study of crystalline solids but extends to include liquids and amorphous materials that have partial ordering. Its primary aim is to provide an explanation of the macroscopic properties of materials through an understanding of underlying atomic structure. For example, we are now able to understand why metals are opaque and are generally good conductors of electricity. Early attempts to understand solid state materials pre-date modern physics and employed empirical knowledge to extract and manipulate materials. An important milestone was passed with the observation of regularities in the morphology of cleaved crystalline structures in the late 18th century. It was not until the early 20th century that the connection between atomic structure and macroscopic appearance was demonstrated through pioneering X-ray diffraction experiments by Max von Laue and the first consistent theory of diffraction by William L. Bragg. This confirmed the belief that solids state materials consist of ordered arrangements of atoms.

Solids state theory has benefited greatly from advances in mathematics and modern physics over the past century. Notable contributions include the application of group theory to the study of crystallographic symmetry systems and quantum theory as the foundation for the description of phenomena manifest in the solids state. Today, solid state physics is considered a mature field having a enormous scope with a significant proportion of physicists engaged in the study in one form or another. This is largely due to the broad spectrum of phenomena that can be observed in the solids phase. These can be categorised as electronic, structural, optical, magnetic, thermal, nuclear and phenomena that involves two or more of the aforementioned.

The field of solid state physics is amongst the most successful. Not only has it answered many important and ancient questions about the nature of materials, it has also given rise to a plethora of applications of economic and anthropological importance. Applications ranging from semiconductor devices found in the modern day computer to the jet engine powered aeroplane were all made possible through the study of solid state materials. The field is also amongst the most diverse, an effect that is inherent in part from the wide range of topics that it encompasses. We can contrast this with disciplines such as nuclear physics that began with the discovery of uranium fission and molecular biology with the discovery of DNA. This is however a testament to the far reaching uses of solid state physics.

Advances in fundamental physics, materials fabrication and analysis tools have led to recent developments in solid state physics. Modern topics that are studied intensively include strongly electronic correlated systems, photonics, photovoltaics, cavity quantum electrodynamics, liquid crystals, low dimensional structures, topological insulators, quasi-crystals, spintronics and nanoscale structures. Such topics can be described as more advanced as they often include electronic many-body effects. For example, in strongly correlated systems, electron-electron correlations dominate at low energies and thus require a more complete treatment to accurately predict the materials properties. The experimental tools for materials characterisation have also benefited from advances in physics as a whole. The development of the laser has made spectroscopic techniques routine in many laboratories worldwide. Moreover, the use of large scale accelerator faculties for materials characterisation has also increased as techniques such as neutron scattering, various X-ray spectroscopic techniques, coherent X-ray imaging and mass spectrometry play an ever increasing role in the study of solid state materials.

Where experiment proves difficult or inaccessible, computational modelling has emerged as a "third pillar" in scientific investigation to explain the various physical properties of materials. Standard tools for materials modelling include density functional theory (DFT) and dynamic mean field theory (DMFT). While often computationally intensive, materials modelling permits investigation of materials that can be tailored at will while predicting new phenomena.

A number of phenomena have proven more difficult to understand and remain a challenge to study using the available experimental tools. Advances here will rely on accurate models that are able to explain data obtained from experiments and predict further useful experiments to perform. Future research in solid state physics will continue to investigate topics including metal-to-insulator phase transitions, magnetic ordering, high temperature superconductivity, topological ordering, structural phase transitions and nanoscale materials science.

Schedule and Assessment

In this course we will focus on the study of crystalline solids and omit discussions on liquids and amorphous materials. These course notes define the content to be delivered for the Level 3 core module on Crystalline Solids. The course consists of a double and single lecture delivered each week for the duration of the first semester. We will in general devote the single lecture slot to problem solving, based partly on non-assessed problem sheets and partly on past exam papers. You are strongly advised to work through all non-assessed problem sheets independently before the lecture.

Assessment consists of a written examination at the end of the course that carries 90 % of the final mark. The exam paper contains two sections. Section A is compulsory and consists of shorter questions potentially taken from all parts of the course. Section B is divided into three or four structured questions, with marks awarded from two questions at most. A series of assessed problem sheets carry the remaining 10 % of the final mark. They will become available in succession during the semester and should be completed before the semester end.

Course material is available on Blackboard, the augmented learning website.

The recommended course book is "**Introductory Solid State Physics**" by **H.P. Myers**.

There are many excellent texts on this subject, mostly somewhat more advanced. The most common are listed below with their strengths and weaknesses:

1. N. W. Ashcroft & N. D. Mermin, Solid State Physics (1976)
   Benefits: Classic text with good depth.
   Weaknesses: Quite mathematical. Generally less accessible.

2. C. Kittel, Introduction to Solid State Physics (2008)
   Benefits: Well structured text.
   Weaknesses: Some topics lack rigour in their treatment. Greater knowledge of quantum mechanics is assumed. This can make the text more challenging to grasp.

3. J. R. Hook and H. E. Hall, Solid State Physics, Second Edition (2010)
   Benefits: Fairly comprehensive. Good coverage of topics.

Weaknesses: Not the most logically ordered text. The wording could be improved in places.

4. H. Ibach and H. Luth, Solid-State Physics (2009)
Benefits: Good coverage of topics.
Weaknesses: Quite light at times when explaining topics.

Marcus C. Newton
Physics & Astronomy
University of Southampton

Contents

Nomenclature

a	Lattice constant
$\mathbf{a}$	Primitive lattice vector
$\mathbf{a}^*$	Primitive reciprocal lattice vector
B	Magnetic field
B_J	Brillouin function
$\mathbf{B}$	Magnetic field vector
C	Curie's constant
d_n	Distance
D	Density of states
D_e	Diffusivity (diffusion constant for electrons)
e	Electron charge
E	Energy
E_g	Energy band gap
$\mathcal{E}$	Electric field magnitude
$\mathcal{E}_{\mathrm{det}}$	Electric field magnitude at detector
$\mathbf{E}$	Electric field vector
g	Reciprocal lattice vector (in one-dimension)
g	Landé splitting factor (chapter 8)
$\mathbf{G}$	Reciprocal lattice vector
$\hbar$	Planck's constant$/2\pi$
h	Miller index
k	Miller index
l	Miller index
H	Hamiltonian
I	Current
j	Current density
$\mathscr{J}$	flux (probability current)
L	Length
k	Wavevector magnitude
k_B	Boltzmann's constant
$\mathbf{k}$	Wavevector
$\mathbf{k}_{\mathrm{in}}$	Incident wavevector
$\mathbf{k}_{\mathrm{out}}$	Reflected wavevector
m	Mass
m_e	Mass of electron
m_n	Mass of neutron
m_h	Mass of hole
m_J	Angular momentum
M	Magnetisation
$\mathbf{M}$	Magnetisation vector
n	Number density
N	Number, integer
N_e	Number of electrons per atom
p	Momentum
$\mathbf{p}$	Momentum vector
$\mathbf{p}$	Dipole moment (chapter 1)

q	Electric charge of particle (e.g.: electron, hole)
Q	Electric charge
$\mathbf{Q}$	Wavevector transfer
r	Distance, separation
$\mathbf{r}$	Position in space
R_H	Hall coefficient
$\mathbf{R}_n$	Lattice vector
ΔS	Change in entropy
$S(\mathbf{Q})$	Structure factor
t	Time
T	Temperature
$u(\mathbf{r})$	Periodic potential
U	Potential Energy
u	Basis vector coefficient
v	Basis vector coefficient
w	Basis vector coefficient
v	Velocity
v_g	Group velocity
v_F	Fermi velocity
v_d	Drift velocity
V	Volume
V	Voltage SI unit (note: not italic.)
$\mathcal{V}$	Electric potential, applied voltage
δ	Dirac delta function
η	Order Parameter
μ	Mobility
ψ	Wave function
ρ_e	Electron density
σ	Conductivity
τ	Relaxation time
χ	Magnetic susceptibility
ω	Natural frequency
ω_p	Plasma frequency

Chapter 1

Bonding in Solids

1.1 Introduction

Solids in general, and in particular crystalline solids, are characterised by a very orderly arrangement of atoms. In the case of crystalline solids the lattice has full translational symmetry at least, i.e. if the crystal is moved in any direction by a lattice vector it looks and can be considered the same as before it was moved. Thus the solid state has very little entropy (is very ordered and thus "unlikely"). Why therefore do solids form? Thermodynamics tells us that if the system's entropy has to decrease to form a solid the system has to give energy to the surroundings otherwise solids would not form. From this experimental evidence we can deduce that solids form because atoms and molecules exert attractive forces on each other. These attractive forces between atoms [1] in solids are called bonds. However we know by measuring the density of a wide range of materials that although atoms are mostly made up of empty space there is a limit to how close they will pack together. That is, at some equilibrium distance the forces of attraction between atoms become zero and for distances closer than that the atoms repel. One way of representing this is to consider the potential energy that two atoms have due to each other at different distances. A generic form of this is given in the diagram shown in figure 1.1, below. The zero of potential energy is chosen to be when the atoms are infinitely far apart. As force is equal to the gradient of potential energy it is clear that the equilibrium separation will be at the minimum of the potential energy. We can also read off from the graph the energy required to separate the atoms to infinity, the binding energy.

One thing that is easily observed is that different materials crystallise at different temperatures. We know from thermodynamics that the temperature at which crystallisation occurs, the melting or sublimation temperature, is reached when the entropy decrease of the material on crystallising is equal to the energy given out due to crystallisation divided by the temperature; $\Delta S = \Delta E / T$. Therefore the melting or sublimation temperature provides us with our first piece of evidence about bonds in solids with which to build our theories. Materials with high melting or sublimation temperatures give out more energy when the bonds are formed i.e. they have stronger bonds. However as crystals often form from liquids, in which the atoms or molecules are relatively close and therefore still interacting, the melting or sublimation temperature is not a perfect measure of the energy stored in the bonds.

From Table 1.1 we can see that there is a huge range of melting/sublimation temperatures.

---

[1] Attractive forces also exist between molecules and their study is central to understanding the liquid state.
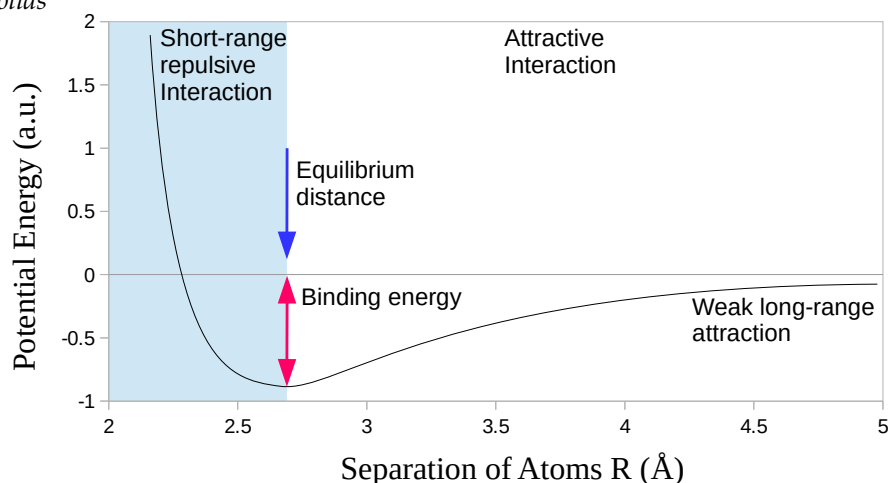
Figure 1.1: A schematic representation of the interaction potential energy between two atoms. The equilibrium separation occurs at the minimum of the potential energy. The depth of the "well" is known as the Binding energy and is deeper for atoms with stronger interaction.

This strongly suggests that the there must be more than one type of bond in solids, i.e. the forces between argon atoms and between silicon atoms are quite different.

What other experimental evidence is available to us to help us understand bonding? We could perform a simple experiment to measure the forces on the atoms by compressing or stretching a material and measuring its elastic constants. Such measurements do indeed show that silicon is much stiffer than ice, as expected from the stronger bonds in silicon, however these mechanical properties of a material often strongly depend on the history of the sample and defects within the sample, and are not simple to understand.

Table 1.1: Melting/sublimation temperature for a number of different solids.

Substance	Melting/Sublimation Temperature (K)
Argon	84
Nitrogen	63
Water	273
Carbon Dioxide	216 - sublimation
Methanol	179
Gold	1337
Salt (NaCl)	1074
Silicon	1683

In X-ray diffraction, which will be described in more detail in later sections, X-ray photons are being scattered from the electrons in a solid. In fact, X-ray diffraction allows us to map out the electron density within solids, i.e. it provides a map of the probability of finding an electron at a certain place within the crystal. Two such maps, for diamond and NaCl, are shown below in figure 1.2. These maps show that most of the electrons in the solid are still held quite close to the nuclei of the atoms; the electron density is much higher nearer the center of the atoms. In addition they show that the orbitals of these electrons are not much disturbed from their positions in atoms in a dilute gas, i.e. they are still spherical. These electrons are often called the core electrons. However in both NaCl and diamond not all the

electrons are undisturbed in the solid compared to the gas. In the case of NaCl the X-ray diffraction shows that one electron has mostly transferred from the Na atom to the Cl atom forming Na^+ and Cl^- ions; this can be seen by integrating the electron density over a sphere centred on each ion to determine the number of electrons associated with each atom. In the case of diamond it appears that some of the electrons from each carbon, actually 4 electrons from each carbon (one for each of its neighbouring carbons), can now be found between the carbon atoms as if they were now being shared between neighbouring carbon atoms. This can be seen from the fact that the electron density between atoms A and B is no lower than 5 however in the areas away from the atoms the electron density drops to 0.34. From these observations we can deduce that although most of the electrons, the core electrons, of an atom are relatively undisturbed in a crystal some of the electrons are quite strongly disturbed.
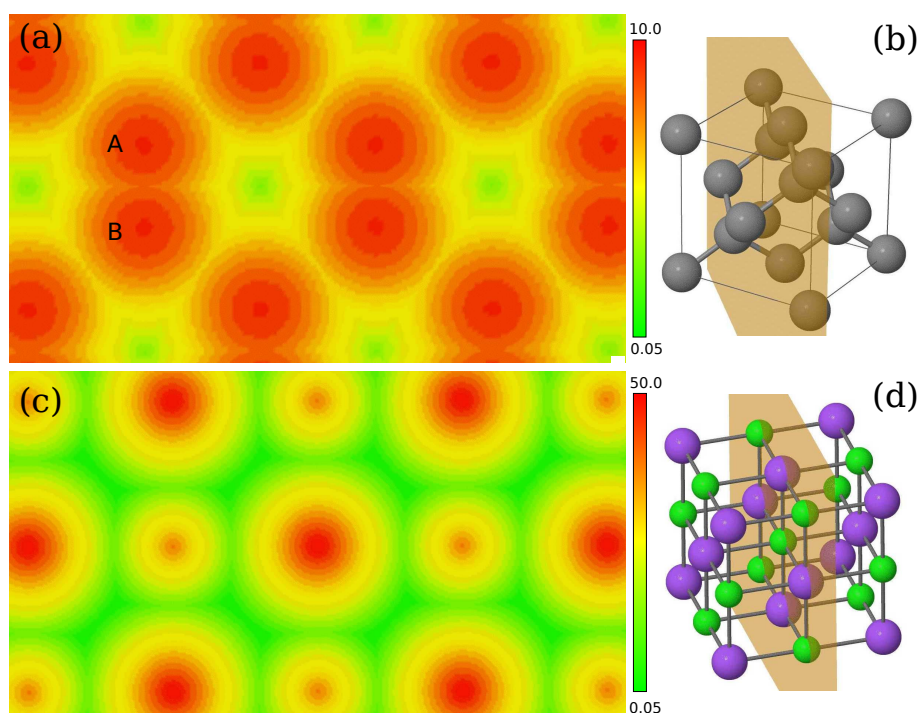
Figure 1.2: (a), X-ray diffraction map of electron density for diamond, shown along the ($4\bar{4}0$) plane. Units of the scale bar are electrons per cubic Angstrom. (b), unit cell of diamond showing the orientation of the plane in (a). (c), electron density map for NaCl showing the (100) plane through one face of the cubic unit cell (d).

We have now gathered quite a lot of information about how solids are bound together and so we will now discuss qualitatively some of the theories that have been developed to understand the process of bonding in solids. Before we go on to discuss the various standard bonding models we first need to answer a very general question. Which of the fundamental forces of nature is responsible for bonding in solids? The answer is the electromagnetic force. Gravity is too weak and the nuclear forces are too short range. The general form of the attraction is shown in the schematic of figure 1.3. The basic concept is that when two atoms are brought together the nuclei of one atom is electromagnetically attracted to the electrons of the other atom. For bonding to occur the total attraction between electrons and nuclei has to be stronger than the repulsion due to the electron-electron and nuclei-nuclei interactions. Although in many cases these attractions are not discussed explicitly this is the underlying mechanism behind all the different forms of bonding in solids.

Figure 1.3: Two lithium atoms in close proximity showing the electromagnetic interactions between the nuclei and electrons. The solid lines denote attractive interactions, the broken lines repulsive interactions.

1.2   Covalent Bonding

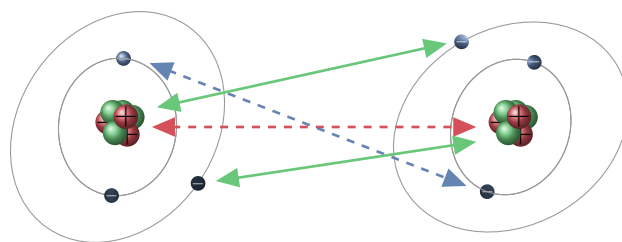Covalent bonding is so named because it involves the sharing of electrons between different atoms with the electrons becoming delocalised over at least two atoms. It is an extremely important form of bonding because it is responsible for the majority of bonding in all organic molecules (carbon based) including nearly all biologically important molecules. In order to try and understand covalent bonding let us first look at covalent bonding in the simplest possible molecule H_2^+ i.e. two protons and one electron. It is clear that if we separate the two nuclei very far apart then the electron will be associated with one nuclei or the other and that if we could force the two nuclei very close together then the electron would orbit the two nuclei as if they were a helium nucleus. What happens for intermediate separations? To answer this we need to solve Schrödinger's wave equations for the electron and two nuclei. This equation contains terms for the kinetic energy of the electron and the two nuclei and all the electrostatic interactions between the three particles. It is in fact impossible to solve this problem analytically as it involves more than two bodies. Therefore we need to make an approximation to make the problem solvable. The standard approximation is called the Born-Oppenheimer approximation and it is based on the fact that the electron is 2000 times less massive than a proton. This means that a proton with the same kinetic energy as an electron is moving roughly 50 times slower than the electron. The Born-Oppenheimer approximation states that as far as the electron is concerned the protons (nuclei) are stationary. We then solve the Schrödinger equation for the electron to obtain the eigenstates (orbitals) and their energies. The groundstate and first excited state electron orbitals obtained from such a calculation for different nuclear separations are shown in Fig. 1.4. The energy of these orbitals as a function of nuclear separation is also shown in Fig. 1.4. We can see from the latter figure that only the lowest energy orbital gives a lower energy than infinitely separated atoms and thus only if the electron is in this orbital will the nuclei stay bound together. For this reason this orbital is called a bonding orbital whereas the first excited state is called an anti-bonding orbital.

What causes the difference between the two states? This can be clearly seen when we examine the electron density along the line joining the two nuclei and compare them against the electrostatic potential energy of the electron due to the two nuclei along the same line. The bonding orbital concentrates electron density between the nuclei where the electrostatic potential energy is most negative. Where as, the anti-bonding orbital has a node half way between the nuclei and its electron density is mostly outside the region where the electrostatic potential energy is most negative.
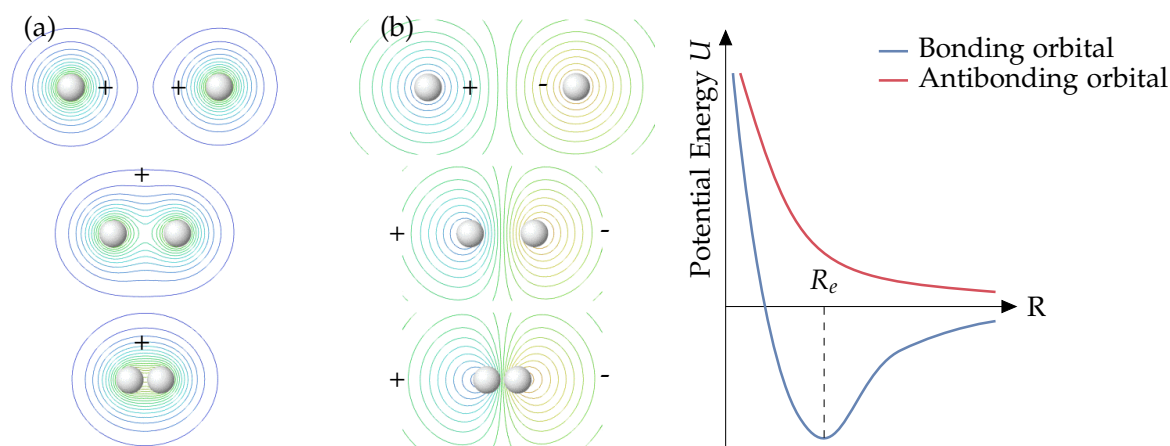
Figure 1.4: Electron density maps for (a) the lowest energy (bonding) and (b) the first excited (anti-bonding) electron orbitals of the molecule H_2^+ at different fixed (Born-Oppenheimer Approximation) nuclear separations R. "$+/-$" labels indicate the sign of the wave function. The graph on the far right shows the energy of the molecule as a function of nuclear separation for the bonding and anti-bonding orbitals.

From Fig. 1.5 we can see that even if the electron is in the bonding orbital that the molecule's energy has a minimum for a finite separation and then increases for smaller separation. From the calculations we can show that for this molecule this is mostly because of the electrostatic repulsion between the nuclei however as we will discuss later other effects can be important in bonds between atoms with many electrons.
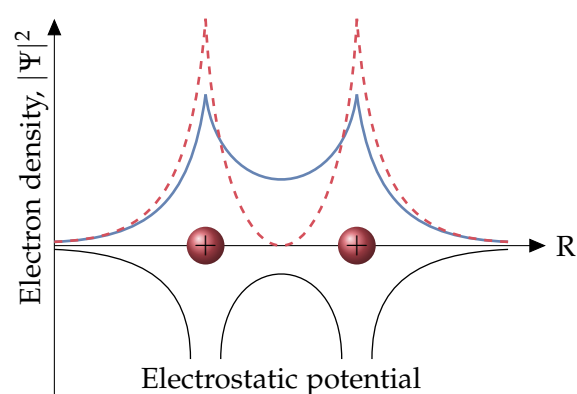
Figure 1.5: Comparison of Bonding and Anti-Bonding orbitals against the electrostatic potential due to the two nuclei.

Even with the Born-Oppenheimer approximation solving for the orbitals of the H_2^+ is complicated and so often we make another approximation which is based on the fact that the true molecular orbitals shown above are well approximated by a sum of two atomic 1S orbitals, one on each atom. This is shown schematically in Fig. 1.6.

This approximation, that the molecular orbitals can be made from the sum of atomic orbitals, is called the linear combination of atomic orbitals[2]

So far we have only treated the simplest of molecules. How do we treat molecules with more than one electron? Well we go about this in a similar manner to how we treat atoms with more than one electron. That is we take the orbitals calculated for a single electron and add electrons one by one taking into account the Pauli exclusion principle. So what happens with H_2 the next most complicated molecule. Well the two electrons can both go into the bonding orbital and H_2 is a stable molecule. What about He_2? Well we have 4 electrons, 2 go into the bonding

---

[2]In fact the method is only approximate because we are only using one atomic orbital for each nuclei. If we were to use all of the possible orbitals the method would be exact.

$$\Psi_{bonding} = (\psi_1 + \psi_2)/\sqrt{2} \qquad\qquad \Psi_{anti\text{-}bonding} = (\psi_1 - \psi_2)/\sqrt{2}$$

Figure 1.6: Schematic representation of the Linear Combination of Atomic Orbitals. The bonding orbital is produced by constructive interference of the wavefunctions in the area between the nuclei and the anti-bonding orbital by destructive interference.

orbital and 2 into the anti-bonding orbital. In fact it turns out that the anti-bonding orbital's energy is greater than two separate atoms 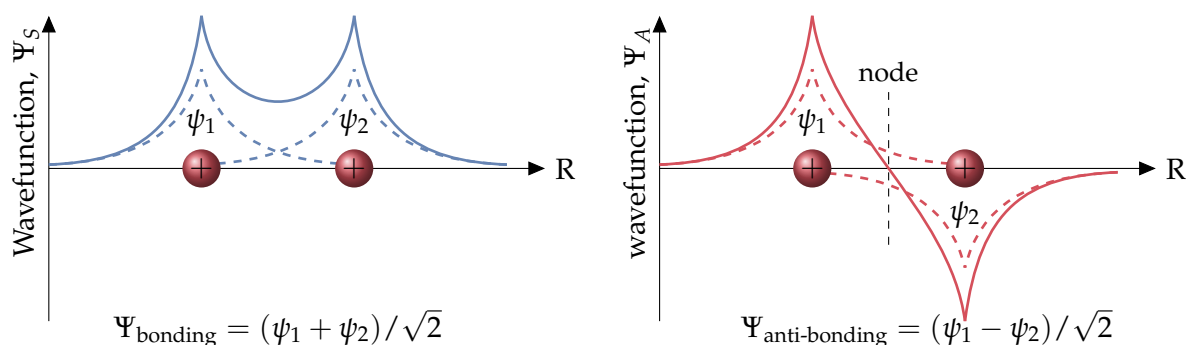by more than the bonding orbital is less than two separate atoms and so the whole molecule is unstable and does not exist in nature. What about Li_2 [3]. In this case we have 6 electrons the first four go as in He_2 and are core electrons. The contribution of these electrons to the total energy of the molecule is positive and leads to a repulsion between the nuclei. For bonding between atoms with more than one electron each, this repulsion adds to the repulsion due to the electrostatic repulsion between nuclei to give the repulsion that keep the atoms from coming too close. The final two electrons go into the bonding orbital formed from the two 2s atomic orbitals. These electrons make a negative contribution to the molecules total energy large enough that Li_2 would be stable. As the 2s orbitals extend further from the nuclei than 1s and the core electrons lead to a repulsive interaction between the nuclei the equilibrium nuclear separation will be relatively large compared with the 1s orbitals. We know from the calculations of molecular orbitals that this means the core orbitals will be very similar to separate 1s orbitals. Which explains why core electrons do not seem to be disturbed much by bonding.

So far we have only discussed molecular orbitals formed from linear combinations of s symmetry atomic orbitals however of course it is possible to form covalent bonds from p, d and all the other atomic orbitals. As the p and d orbitals can be thought of as pointing in a certain direction from the atom (Fig. 1.7), it is these orbitals that produce the directionality in covalent bonding that leads to the very non close packed structure of many covalent solids such as diamond. However this subject is more complex and I will not deal with it explicitly. If you are interested I suggest you read more in a basic chemistry textbook like Atkins (Chapter 10).

Figure 1.7: Constant electron density surface maps of s, p and d atomic orbitals

In addition, we have also only discussed covalent bonding between identical atoms. If the two atoms are not identical then instead of the electrons in the bond being shared equally between the two atoms the electrons in the bonding orbitals will spend more time on one of the atoms
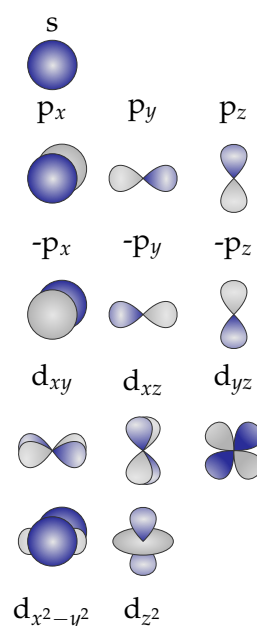
---

[3]Of course lithium is a metal and doesn't naturally form diatomic molecules.

and those in the anti-bonding orbital on the other atom. However for atoms that are fairly similar like carbon, oxygen, sulphur etc. the split of the electrons between the two atoms is fairly equal and the bonds are still mostly covalent in nature. If the atoms are very different like Na and Cl then we have instead ionic bonding.

1.3 Ionic Bonding

When discussing covalent bonds between two different atoms I talked about similar and different. These are not very specific terms. Why are carbon and sulphur similar and sodium and chlorine different. In terms of ionic and covalent bonding there are two energies that are important for determining whether the atoms form an ionic or covalent bond[4]. These are the electron affinity, the energy difference between a charge neutral atom or negative ion with and without an extra electron (e.g. $O \rightarrow O^-$), and the ionisation energy, the energy required to remove an electron from the a charge neutral atom or positive ion, (e.g. $O \rightarrow O^+$). In Table 1.2 we present these energies for the series of atoms Li to Ne. When we compare the magnitude of the electron affinities and first ionisation energies to the atomic radii of each neutral atom, as shown in figure 1.8, a clear trend becomes apparent. It is left to the reader to think about why the series show the trends that they do.

Table 1.2: Electron affinities and first ionisation energies for the second row of the periodic table

Element	First Ionisation Energy (eV/atom)	Electron Affinity (eV/atom)
Lithium	5.4	-0.62
Beryllium	9.3	No stable negative ions
Boron	8.3	-0.28
Carbon	11.2	-1.27
Nitrogen	14.5	0
Oxygen	13.6	-1.43
Fluorine	17.4	-3.41
Neon	121.5	No stable negative ion

Now imagine we were to remove an electron from a carbon atom and give it to another carbon atom the total energy required to make this change is 11.2 - 1.27 = 9.93 eV/atom. However if we were instead to take an electron from a lithium atom, forming a Li^+ ion, and give it to a fluorine atom, forming a F^- ion, the total energy required would be 2 eV/atom, which is much less. The energy we have just calculated is for the case that the atoms, and ions which form from them, are infinitely far apart. However if we were to bring the ions together then the total energy of the system would be lowered. For Li^+ and F^- ions 0.35 nm apart, the separation in a LiF crystal, the electrostatic potential energy would be 4 eV. Of course in a LiF crystal the Li^+ ions don't interact with a single F^- but with all the other ions in the crystal, both Li^+ and F^-.

---

[4]Note that although pure covalent bonds can occur between two identical atoms pure ionic bonding is not possible as the electrons on a negative ion will always be attracted to towards positive ions and therefore be more likely to be found in the volume between the ions i.e. be shared. Thus ionic and covalent bonds should be considered ideals at either end of an experimentally observed continuum.
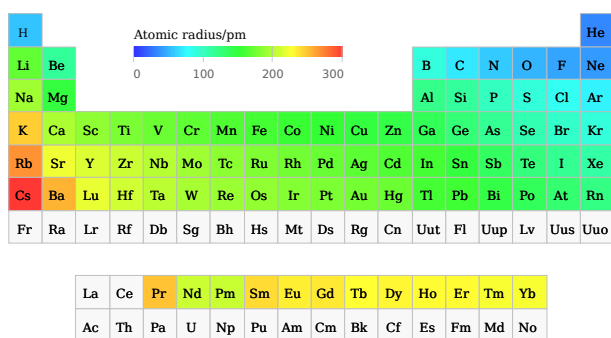
Figure 1.8: Atomic radii (in picometers) of s, p and d-block elements of the periodic table.

When all of these interactions are taken into consideration the magnitude of the electrostatic potential energy of the crystal is much larger than the energy cost for forming the ions and therefore the crystal is stable. This type of bonding where electrons transfer to form ions that are electrostatically bound is called ionic bonding[5].

We have only discussed the transfer of one electron between atoms however a large number of ionic crystals do exist with ions with more than one excess or removed electron. For instance it is relatively easy, i.e. low energy cost, to remove two electrons from group II atoms and group VI atoms will easily accept two electrons. Also there are a number of ions that are made up of more than one atom covalently bound, for instance CO_3^{2-}.

1.4   Metallic Bonding

This bonding is in many ways similar to the covalent bond in as much as the electrons are shared amongst the atoms. It's main difference is that in this type of bonding the electrons are not localised in the region between atoms but instead are delocalised over the whole crystal. Another way of looking at metallic bonding is that every atom in the metal gives up one or more electrons to become a positive ion and that these ions are sitting in a "sea" or "cloud" of free electrons. The atoms stay bound because the electrostatic interaction of the electrons with the ions.
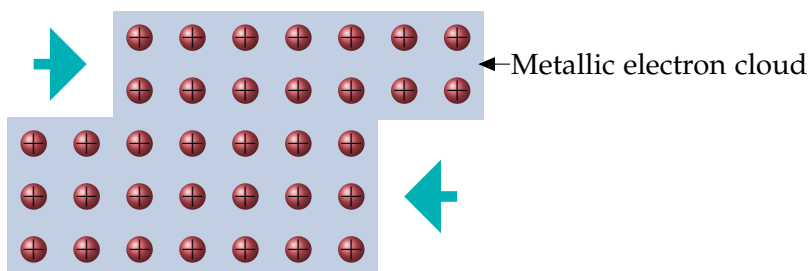
Figure 1.9: Schematic representation of a metallically bonded material. Charged ions surrounded by a "cloud" or "sea" of delocalised electrons. Layers of atoms are able to slide over each other when a force is applied to the metal.

The bonds between atoms in a metal are strong enough to maintain cohesion but are generally not rigid. As a result, a sufficiently large force applied to a metal can cause layers of atoms to 'slide' across one another. After this slipping event has occurred the atoms tend to settle into position again and an equivalent structure is restored. It is this effect that endows metals with both ductile and malleable properties.

Now is a convenient time to stress that the bonding models we have been discussing are

---

[5]In fact pure ionic bonds do not exist, unlike pure covalent bonds, as the electrons around any negative ions are attracted to its positive neighbours they are more often found between the ions and all ionic bonds have a very small covalent component.

theoretical concepts and that often in real materials the bonding is a mixture of a number of different models. A good example of this is the way that it is possible to describe the metallic bond as being like a covalent and an ionic bond. The key to understanding metallic bonding is that the electrons are delocalised.

1.5   Van der Waals Bonding

When discussing covalent bonding we stated that He_2 is not stable and for the same reason we would not expect Ar_2 to be stable, however Argon does crystallise. What is the bonding mechanism in this case. It can't be ionic as we don't have two types of atoms. We also know that solid argon is not metallic[6]. In fact it turns out that in solid argon the bonding mechanism is a new type called van der Waals' bonding. Argon atoms, like all atoms, consist of a positive nucleus with electrons orbiting it. Whenever an electron is not at the center of the nucleus, it and the nucleus form a charge dipole. Now as the electron is constantly moving the average dipole due to the electron and nucleus is zero[7]. However let us consider an atom frozen in time when the electron and nucleus form a charge dipole $\mathbf{p}_1$. A charge dipole will produce an electric field, $\mathbf{E}$, of magnitude proportional to $\mathbf{p}_1$. This electric field can electrically polarise neighbouring atoms, i.e. cause them to become dipoles aligned parallel to the electric field due to the first atom's dipole. The magnitude of the dipole in the second atom, $\mathbf{p}_2$, will be proportional to the electric field, $\mathbf{E}$. Now a dipole in the presence of an electric field has an electrostatic energy associated with it given by $\mathbf{p} \cdot \mathbf{E}$. Thus, the dipole at the second atom will therefore have an energy due to being in the presence of the electric field due to the first atoms dipole as $\mathbf{p} \cdot \mathbf{E}$. As $\mathbf{p} = \alpha \mathbf{E}$ the energy of the second dipole due to the first is proportional is the magnitude of the first dipole squared i.e. $\alpha |\mathbf{p}|^2$. Now although the average, over time, of the first atoms dipole is zero the interaction energy we have just derived does not depend on the direction of the dipole of the first atom therefore it will not average to zero over time. This fluctuating dipole- fluctuating dipole interaction between atoms is the van der Waals' interaction. From the fact that the electric field due to a dipole falls off as $1/r^3$ we can deduce that the van der Waals' interaction falls off as $1/r^6$.

1.6   Hydrogen Bonding

Hydrogen bonding is extremely important for biological systems and is necessary for life to exist. It is also responsible for the unusual properties of water. It is of intermediate strength between covalent and van der Waals bonding. It occurs between molecules that have hydrogen atoms bonded to atoms, such as oxygen, nitrogen and chlorine, which have large electron affinities. Hydrogen bonding is due to the fact that in the covalent bond between the hydrogen and the other atom the electrons reside mostly on the other atom due to its strong electron affinity (see Fig. 1.10). This leads to a large static electric dipole associated with the bond and this will attract other static dipoles as set out above for van der Waals bonding.

---

[6]At least not at atmospheric pressure.

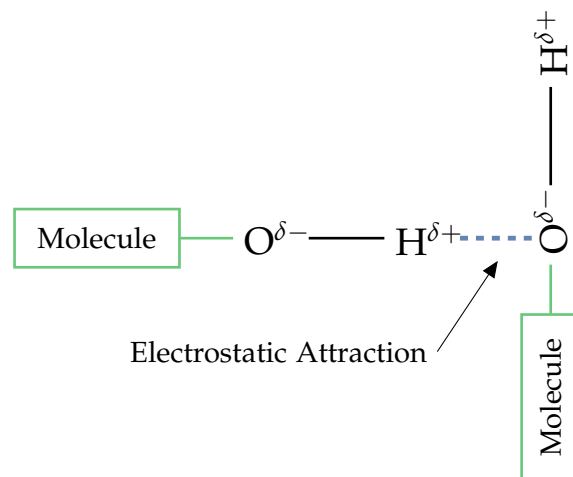[7]Assuming the atom is not in an external electric field

Figure 1.10: Model of hydrogen bond forming between two permanent dipoles.

Chapter 2

Crystal Lattices

2.1   The Crystal Lattice

Although some materials are disordered, many can be understood from the fact that the atoms are periodically arranged. So instead of thinking about 10^{23} atoms we can think about two or three, together with the underlying symmetry of the crystal. We are going to ignore for the moment that crystals are not infinite and they have defects in them, because we can add these back in later.
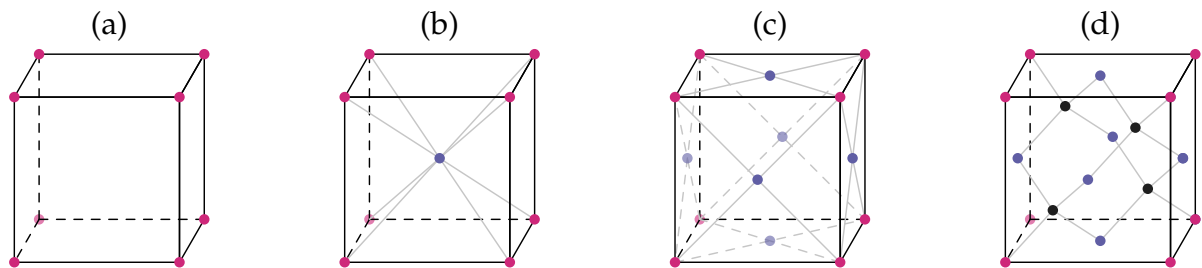
Figure 2.1: Common crystal structures; (a) simple cubic [Iron sulphide] {filling fraction 0.52} , (b) body centred cubic [lithium, molybdenum] {filling fraction 0.68} (c) face centred cubic [calcium, silver, gold] {filling fraction 0.74} and (d) diamond cubic [diamond, silicon, germanium] {filling fraction 0.34.}

The first thing we notice is that crystals come in many different forms, reflecting different possibilities for repeated arrangements of the atoms. For instance, gold and silver atoms coalesce into face-centred cubic structure while lithium and molybdenum atoms coalesce into the body-centred cubic structure (figure 2.1). These structures are discussed in more detail below.

If as a starting assumption we assume that solids are made up of the basic blocks, atoms or molecules, which are treated as incompressible spheres, stacked up to form a solid and that the attraction between the units is isotropic[1] we should be able to determine how the units should be arranged by determining the arrangement which gives the largest energy of crystallisation. It is obvious that what we want to do is to obtain the arrangement of the atoms that minimises the separation between atoms and therefore the space not filled up by atoms.

---

[1]Isotropic - the same in all directions.

Such a structure is called close packed. The filling fraction is the ratio of the volume of the atoms in a crystal to the volume of the crystal. The filling fractions for the structures in Figure 2.1 are given in the caption. It is clear from this that many crystals don't form in a structure that minimises the empty space in the unit cell i.e. they are not close packed. This leads us to review our assumptions as clearly one is incorrect. In fact it turns out that in some materials the attraction between units is directional. Isotropic or directional attraction is one important way of differentiating between bonding in different materials.

In this course we are not going to attempt any sort of exhaustive survey of crystal structures. The reason is that there are an incredible number although they can be fairly well classified into different 'classes'. What we will do is introduce a formalism because it will help to view the diffraction of lattices in a powerful way, which will eventually help us understand how electrons move around inside the lattices. The first part of our formalism is to divide the structure up into the **lattice** and the **basis**.

2.1.1 The lattice

The *lattice* is an infinite **array of points** in space arranged so that each point has identical surroundings. In other words, you can go to any lattice point and it will have the same environment as any other. Each atom in a crystal is necessarily associated with a single lattice point but is not required to sit *at* that lattice point - it can be situated at some fixed distance, as we shall see in the following section. Nearest-neighbour lattice points are separated from each other by the 3 fundamental translation vectors, known as **primitive lattice vectors**, that define the particular three-dimensional (3D) lattice, which we generally write as $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$. We can reach any point on the lattice by a lattice vector translation such that: $\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$, where n_1, n_2, n_3 are integers. Each lattice point is therefore defined by an integer sums of the set of linearly independent **primitive lattice vectors**, $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$. The choice of lattice vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ is however not unique. If the lattice vectors are chosen to connect points that are not nearest-neighbours, the vectors are no longer considered primitive. In the 2D example of figure 2.2, $\mathbf{R}_1 = 2\mathbf{a}_1 + \mathbf{a}_2 = 3\mathbf{a}_1' + \mathbf{a}_2'$ is a lattice vector.
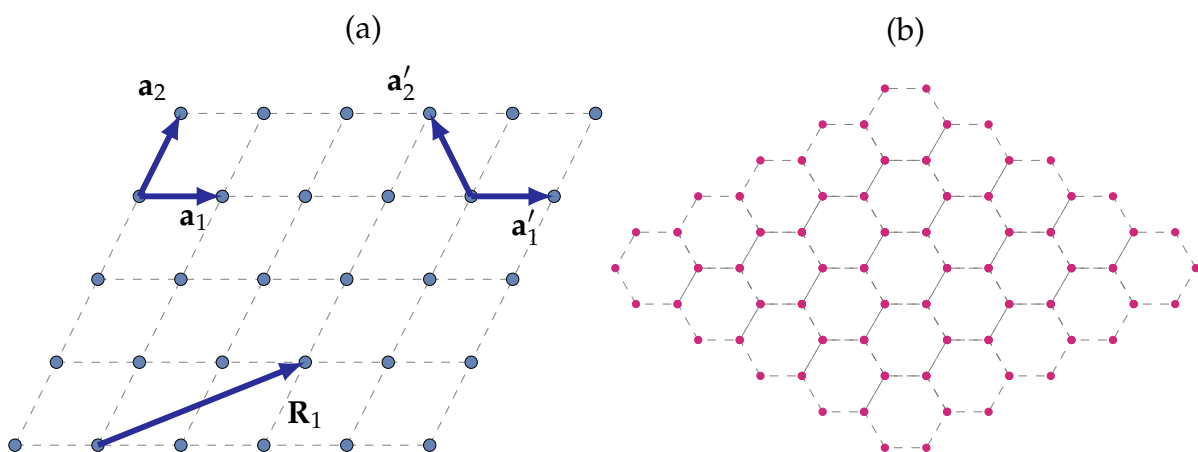

Figure 2.2: (a) Simple 2D lattice structure. Both $\mathbf{a}$ and $\mathbf{a}'$ are primitive lattice vectors while $\mathbf{R}_1$ is not. (b) Honeycomb structure.

Not all regular arrangements of points constitute a lattice. A close look at the honeycomb structure shown in figure 2.2 reveals that each point does not have identical surroundings.

Two types of points are apparent and suggest that the concept of a lattice must be extended to reproduce this structure.

2.1.2   The Basis

The *basis* of the crystal is the **specific arrangement of atoms** associated with each lattice point. So for instance we can have a three atom basis attached to the above lattice to give the crystal structure:
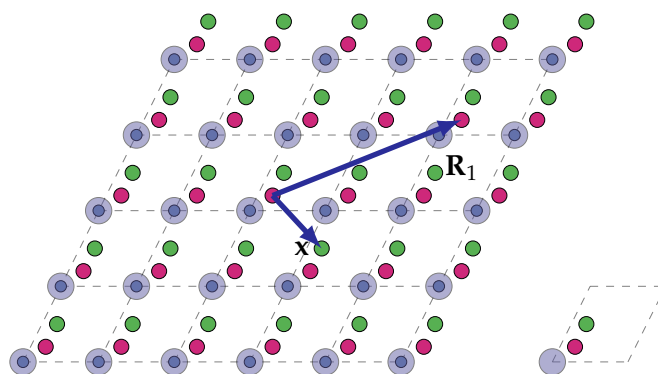

Figure 2.3: Simple 2D lattice structure with basis (lower right).

Moving by any lattice vector (e.g. $\mathbf{R}_1$) goes to a position which looks identical. This reflects the translational symmetry of the crystal. Note that $\mathbf{x}$ is not a lattice vector, as the shift it produces results in a different atomic environment. Most elements have a small number of atoms in the basis; for instance silicon has two. However α-Mn has a basis with 29 atoms! Compounds must have a basis with more than one atom, e.g. CaF_2 has one Ca atom and two F atoms in its basis. Molecular crystals (such as proteins) have a large molecule as each basis, and this basis is often what is of interest - i.e. it is the structure of each molecule that is wanted.

2.1.3   The Unit Cell

Lattice vectors also define the edges of the unit cell (where the associated lattice points define the vertices). We often talk about the **primitive unit cell** of a lattice, which is the **volume** associated with a **single lattice point**. It is also the smallest possible and the only *true* unit cell as it is constructed with primitive lattice vectors. Its name is derived from the fact that we are able to stack such cells together and generate the entire lattice structure.

An alternative means of constructing the primitive unit cell is to consider the unit of mesh formed by connecting planes (lines in 2D) that are perpendicular bisectors of lines that join a lattice point with its nearest-neighbours. The primitive unit cell obtained with this approach results in a lattice point confined to its centre. It is obtained using the following procedure (Figure 2.4):

1. Draw in all lattice vectors from one lattice point

2. Construct all planes which divide the vectors in half at right angles

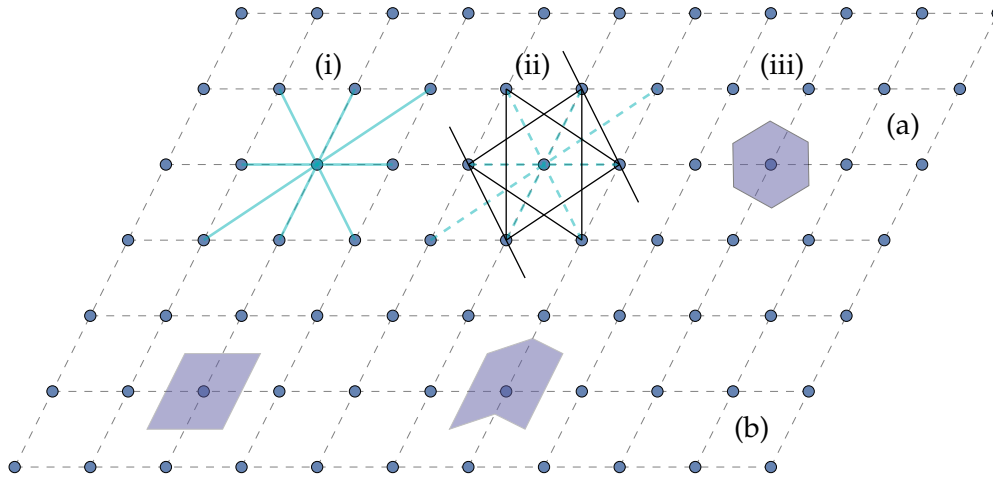3. Find the volume around the lattice point enclosed by these planes

Figure 2.4: Construction of the Wigner-Seitz cell (a), other primitive cells (b).

This type of primitive cell is called a Wigner-Seitz cell, but alternatives are possible with the same volume (or area in this 2D case), and which can be tiled to fill up all of space.

2.1.4 Lattice Systems

On the basis of symmetry, we can catalogue lattices into a finite number of different types ('Bravais' lattices). There are 5 such lattices in two-dimensions and 14 in three-dimensions. In two-dimensions, they are the square, rectangular, oblique, centred-rectangular and hexagonal. Symmetry relations for all of the 14 lattices in three-dimensions are depicted below in table 2.1. Each of these has a particular set of rotation and mirror symmetry axes. The least symmetrical 3D lattice is called triclinic, with the fundamental vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ of different lengths and at different angles to each other. The most symmetrical is the cubic, with $|\mathbf{a}_1| = |\mathbf{a}_2| = |\mathbf{a}_3|$ all at right-angles.

Table 2.1: Symmetry relations for the 14 Bravais lattices.

Basis vectors	Angles	Crystal system
$\mathbf{a}_1 = \mathbf{a}_2 = \mathbf{a}_3$	$\alpha = \beta = \gamma = 90°$	cubic (primitive, bcc, fcc)
$\mathbf{a}_1 = \mathbf{a}_2 = \mathbf{a}_3$	$\alpha = \beta = \gamma \neq 90°$	rhombohedral
$\mathbf{a}_1 = \mathbf{a}_2 \neq \mathbf{a}_3$	$\alpha = \beta = 90°, \gamma = 120°$	hexagonal
$\mathbf{a}_1 = \mathbf{a}_2 \neq \mathbf{a}_3$	$\alpha = \beta = \gamma = 90°$	tetragonal (primitive, body centred)
$\mathbf{a}_1 \neq \mathbf{a}_2 \neq \mathbf{a}_3$	$\alpha = \beta = \gamma = 90°$	orthorhombic (primitive, base / body / face centred)
$\mathbf{a}_1 \neq \mathbf{a}_2 \neq \mathbf{a}_3$	$\alpha = \gamma = 90°, \beta \neq 90°$	monoclinic (primitive, base centred)
$\mathbf{a}_1 \neq \mathbf{a}_2 \neq \mathbf{a}_3$	$\alpha \neq \beta \neq \gamma \neq 90°$	triclinic

2.2 Lattice Positions, Planes and Directions

Positions in the unit cell are defined in terms of the cubic "basis" vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$. The position $u\mathbf{a}_1 + v\mathbf{a}_2 + w\mathbf{a}_3$ is described as *uvw*. Only in the cubic system are these like Cartesian coordinates. u, v and w take values on the unit interval [0,1]. For example, for the:

1. simple cubic lattice there is just one atom at 000

2. body centre cubic lattice (BCC) the atomic positions are 000, $\frac{1}{2}\frac{1}{2}\frac{1}{2}$

3. face centre cubic lattice (FCC) the atomic positions are 000, $0\frac{1}{2}\frac{1}{2}$, $\frac{1}{2}0\frac{1}{2}$, $\frac{1}{2}\frac{1}{2}0$

Directions in the crystal are defined by integer multiples of the fundamental vectors.

The direction of the vector $n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$ is written with 'square' brackets as $[n_1n_2n_3]$. So the body diagonal direction is $[111]$. Negative components are written with a bar above the number instead of a confusing minus sign, e.g. $[1\bar{1}0]$ is a face diagonal.

Planes are written with parentheses as $(n_1n_2n_3)$ - i.e. with round brackets - denoting all the lattice planes orthogonal to a direction $[n_1n_2n_3]$ in the lattice. Planes in the crystal are important for determining the diffraction that can occur. Although Bragg diffraction can be described in terms of the separation of different crystal planes, it is important to note that there are rather a lot of planes in even a sample crystal (Fig. 2.5). Conversely using x-ray diffraction to measure all the spacings allowed by Braggs law does not easily allow you to work back to find the crystal structure itself. You can see this just by trying to draw the planes present in a cubic lattice:
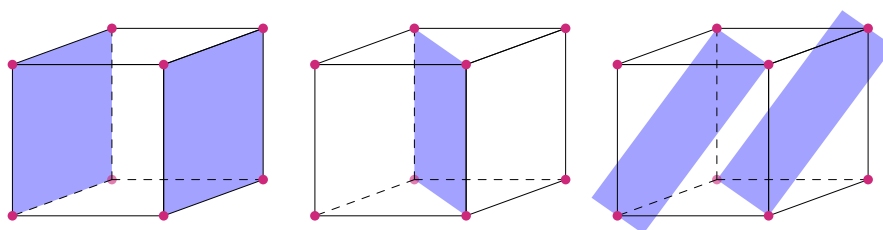
Figure 2.5: Selected planes in the cubic lattice.

This problem becomes much more difficult when we understand that virtually no material has the simple cubic lattice, but have more complicated atomic arrangements. In order to interpret diffraction patterns and understand crystal structures a concept called the "reciprocal lattice" is used which is described in the next section.

2.3   The Cubic Crystal System

In this course we will often use examples of the cubic type which has three common but different varieties: simple cubic, body-centred cubic (BCC), face-centred cubic (FCC) and diamond cubic. The BCC and FCC structures are formed by adding extra lattice points to the unit cell either in the center of the cell, or at the center of each face. In these cases, the cubic unit cell is then no longer primitive, but is still most convenient to work with. This naturally leads to the concept of the **conventional unit cell** with vectors that form the edges of the simple cubic lattice unit cell structure. It should be noted that unlike the honeycomb structure discussed above, the simple cubic, BCC and FCC structures are indeed lattices and can be described with a single atom in the basis.[2] The primitive unit cell of these structures however is not a regular cube and is less convenient for further studies. In the following, we show how each cubic system is described in further detail by making use of the conventional unit cell.

---

[2]The diamond cubic structure is not a lattice. Try to convince yourself why this is the case.

2.3.1 Body Centred Cubic

The body centred cubic lattice (BCC) consists of a simple cubic lattice with 2 atoms in its basis. BCC basis points are located at 000 and $\frac{1}{2}\frac{1}{2}\frac{1}{2}$. The primitive cell (shaded region below) has half the volume of the conventional unit cell. The primitive lattice vectors are then given by:
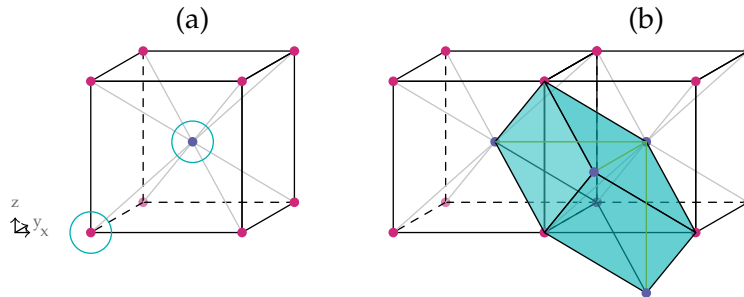
Figure 2.6: BCC lattice structure showing the conventional and primitive (shaded) unit cells. Only two atoms in the basis (highlighted) are necessary to reproduce the entire structure in the simple cubic lattice.

$$\mathbf{a}_1 = [a, 0, 0] \; ; \; \mathbf{a}_2 = [0, a, 0] \; ; \; \mathbf{a}_3 = \left[\frac{a}{2}, \frac{a}{2}, \frac{a}{2}\right] \tag{2.1}$$

2.3.2 Face Centred Cubic

The face centred cubic lattice (FCC) consists of a simple cubic lattice with 4 atoms in its basis. FCC basis points are located at 000, $0\frac{1}{2}\frac{1}{2}$, $\frac{1}{2}0\frac{1}{2}$, $\frac{1}{2}\frac{1}{2}0$. The primitive cell (shaded region below) has one quarter the volume of the conventional unit cell. The primitive lattice vectors are then given by:
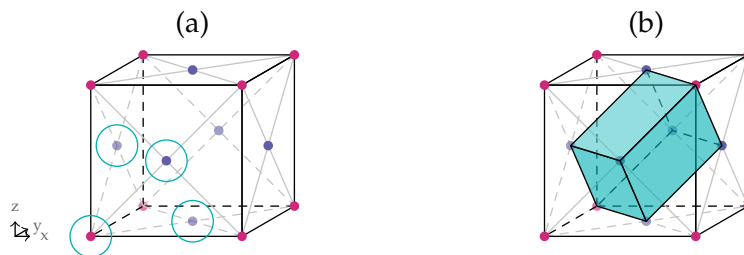
Figure 2.7: FCC lattice structure showing the conventional and primitive (shaded) unit cells. Four atoms are in the basis (highlighted) and are necessary to reproduce the entire structure in the simple cubic lattice.

$$\mathbf{a}_1 = [\frac{a}{2}, \frac{a}{2}, 0] \; ; \; \mathbf{a}_2 = [\frac{a}{2}, 0, \frac{a}{2}] \; ; \; \mathbf{a}_3 = \left[0, \frac{a}{2}, \frac{a}{2}\right] \tag{2.2}$$

2.3.3 Diamond Cubic

The diamond cubic structure consists of a simple cubic lattice with 8 atoms in its basis. These are located at the FCC points with the addition of four points located at a quarter distance

along each of the diagonals (Black dots in figure 2.8). We can also think of the diamond structure as two inter-penetrating FCC sub-lattices, with a basis of 2 atoms situated at the origin 000 position and at $\frac{1}{4}\frac{1}{4}\frac{1}{4}$ along the body diagonal. These two lattices therefore share primitive lattice vectors with the FCC structure.
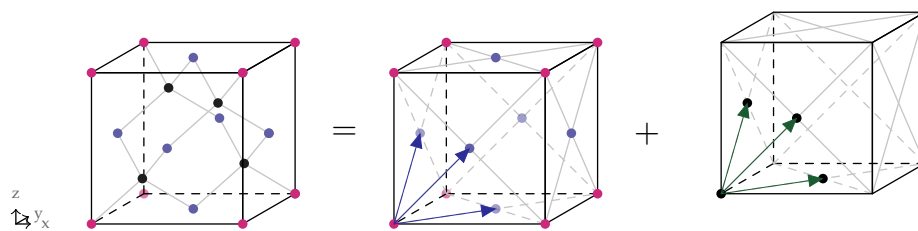

Figure 2.8: Diamond cubic structure showing the conventional unit cell. Eight atoms are in the basis and are necessary to reproduce the entire structure in the simple cubic lattice.

Chapter 3

The Reciprocal Lattice

3.1 Diffraction of Waves from Crystals

In an X-ray diffraction experiment we illuminate a sample with X-rays from a specific direction, i.e. the X-rays are a plane wave, and then measure the intensity of X-rays diffracted into all the other directions. The reason that the crystal causes the X-rays to diffract is that the X-rays electromagnetic field makes the electrons within the material oscillate. The oscillating electrons then re-emit X-rays as a point source into all directions (i.e. Huygens principle in optics). As the phase of the oscillation of all the electrons are set by a single incoming wave the emission from all the electrons interferes to produce a diffraction pattern rather than just a uniform scattering into all directions equally. We will now use Huygens' Principle to calculate the diffraction pattern one would expect from a crystal. Let's start with a lattice of single electrons (see Fig. 3.1 below):
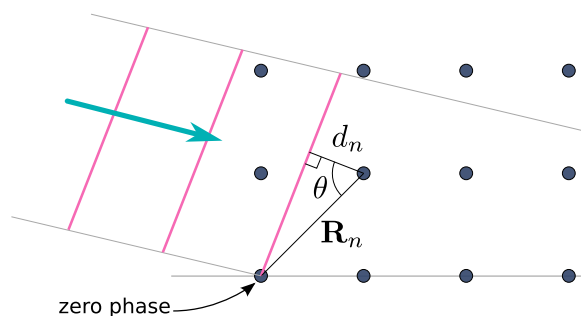
Figure 3.1: X-ray diffraction from a crystal: Incident wave.

As we always have to talk about a relative phase difference we start by defining the phase at one of the electrons to be zero phase. From the diagram we can see that the phase at the electron at position $\mathbf{R}_n$ is given by $2\pi d_n / \lambda$. Standard trigonometry gives that d_n is equal to $\mathbf{R}_n \cos \theta$ i.e. that the relative phase is $2\pi \mathbf{R}_n \cos(\theta)/\lambda$. This expression is more commonly written $\mathbf{k}_{\text{in}} \cdot \mathbf{R}_n$ which is the dot product of the position of the atom and the wavevector of the x-rays. In addition to the different phase of the oscillations of the different electrons we have to take into account the different paths for the emitted x-rays to travel to the detector (see Fig. 3.2 below).

If we set the phase, at the detector, of the X-rays emitted by the zero phase electron to be zero
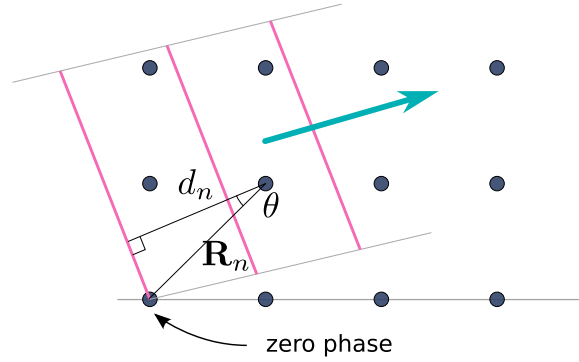
Figure 3.2: X-ray diffraction from a crystal: Diffracted wave.

then the phase, at the detector, of the X-rays emitted by the atom at $\mathbf{R}_n$ will be $-2\pi d_n/\lambda = -\mathbf{k}_{\text{out}} \cdot \mathbf{R}_n$.

Thus the total phase difference of light scattered from the incoming wave at the detector by the atom at $\mathbf{R}_n$ will be $(\mathbf{k}_{\text{in}} - \mathbf{k}_{\text{out}}) \cdot \mathbf{R}_n$ and the total electric field at the detector will be proportional to:

$$\mathcal{E}_{\text{det}} \propto \sum_n e^{i(\mathbf{k}_{\text{in}}-\mathbf{k}_{\text{out}})\cdot\mathbf{R}_n} \tag{3.1}$$

We know that in a real crystal the electrons are not localised and instead should be described by a electron density function, e.g. $\rho_e(\mathbf{r})$. In this realistic case the expression is given by:

$$\mathcal{E}_{\text{det}} \propto \iiint\limits_{\text{crystal}} \rho_e(\mathbf{r})\, e^{i(\mathbf{k}_{\text{in}}-\mathbf{k}_{\text{out}})\cdot\mathbf{r}} dV \tag{3.2}$$

3.2   The Effect of the Lattice

The defining property of a crystal is that it has translational symmetry. The importance of this to X-ray diffraction is that the electron density at a point $\mathbf{r}$ and another point $\mathbf{r} + \mathbf{R}_n$ where $\mathbf{R}_n$ is a lattice vector must be the same. This means that if we write the integral above as a sum of integrals over each unit cell, i.e.:

$$\mathcal{E}_{\text{det}} \propto \sum_{\mathbf{R}_n} \iiint\limits_{\text{UnitCell}} \rho_e(\mathbf{r}+\mathbf{R}_n)\, e^{i(\mathbf{k}_{\text{in}}-\mathbf{k}_{\text{out}})\cdot(\mathbf{r}+\mathbf{R}_n)} dV \tag{3.3}$$

which can be written as:

$$\mathcal{E}_{\text{det}} \propto \left(\sum_{\mathbf{R}_n} e^{i(\mathbf{k}_{\text{in}}-\mathbf{k}_{\text{out}})\cdot\mathbf{R}_n} \right) \iiint\limits_{\text{UnitCell}} \rho_e(\mathbf{r}+\mathbf{R}_n)\, e^{i(\mathbf{k}_{\text{in}}-\mathbf{k}_{\text{out}})\cdot\mathbf{r}} dV \tag{3.4}$$

i.e. the diffracted field has two components. The first term is a sum over the lattice vectors and the second term is due to the charge density in the unit cell. We will first investigate

the effect of the first term. As $\mathbf{R}_n$ is a lattice vector we know it can be written in the form $\mathbf{R}_n = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_2\mathbf{a}_3$. If we define $\mathbf{Q} = \mathbf{k}_{\text{out}} - \mathbf{k}_{\text{in}}$ and substitute for $\mathbf{R}_n$ we get:

$$\mathcal{E}_{\text{det}} \propto \sum_{n_1,n_2,n_3} e^{(-i\mathbf{Q}\cdot(n_1\mathbf{a}_1+n_2\mathbf{a}_2+n_2\mathbf{a}_3))} = \sum_{n_1} e^{(-in_1\mathbf{Q}\cdot\mathbf{a}_1)} \times \sum_{n_2} e^{(-in_2\mathbf{Q}\cdot\mathbf{a}_2)} \times \sum_{n_3} e^{(-in_3\mathbf{Q}\cdot\mathbf{a}_3)} \tag{3.5}$$

Now each of the factors in the multiplication can be represented as the sum of a series of vectors in the complex plane (Argand diagram) where the angle of the vectors relative to the real (x) axis increase from one vector in the series to the next by $\mathbf{Q} \cdot \mathbf{a}_i$. In the general case the sum can be represented by a diagram where the vectors end up curving back on themselves (as below) leading to a sum which is zero.
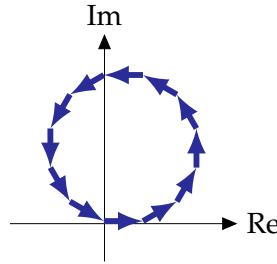
Figure 3.3: Vectors that sum to zero.

The only time when this is not true is when the angle between vectors in the series is zero or an integer multiple of 2π. In this case all of the vectors are along a single line and the sum has magnitude equal to the number of vectors. Whilst the above is not a mathematical proof the conclusion that the scattering is only non-zero in the case that:

$$\begin{aligned} \mathbf{Q} \cdot \mathbf{a}_1 &= 2\pi \times \text{ integer} \\ \mathbf{Q} \cdot \mathbf{a}_2 &= 2\pi \times \text{ integer} \\ \mathbf{Q} \cdot \mathbf{a}_3 &= 2\pi \times \text{ integer} \end{aligned} \tag{3.6}$$

can be shown to be true mathematically. From these expressions we can determine that the X-rays will only be diffracted into a finite set of directions ($\mathbf{Q}$). We can most easily characterise each possible direction by the 3 integers in the above expression which are usually labelled h, k, l. Let us consider the vector difference $\mathbf{a}_1^*$ between the $\mathbf{Q}$ corresponding to h, k, l and that corresponding to $h + 1$, k, l. This vector is defined by the equations:

$$\begin{aligned} \mathbf{a}_1^* \cdot \mathbf{a}_1 &= 2\pi \\ \mathbf{a}_1^* \cdot \mathbf{a}_2 &= 0 \\ \mathbf{a}_1^* \cdot \mathbf{a}_3 &= 0 \end{aligned} \tag{3.7}$$

that is $\mathbf{a}_1^*$ is a vector which is orthogonal to $\mathbf{a}_2$ and $\mathbf{a}_3$ and has a component in the direction of $\mathbf{a}_1$ which is equal to $2\pi/\mathbf{a}_1$.

If we define similar vectors $\mathbf{a}_2^*$ and $\mathbf{a}_3^*$ then the diffraction vector $\mathbf{Q}$ corresponding to the indices h, k, l is given by:

$$\begin{aligned} \mathbf{Q} &= \mathbf{G} \\ \mathbf{G} &\equiv h\mathbf{a}_1^* + k\mathbf{a}_2^* + l\mathbf{a}_3^* \end{aligned} \tag{3.8}$$
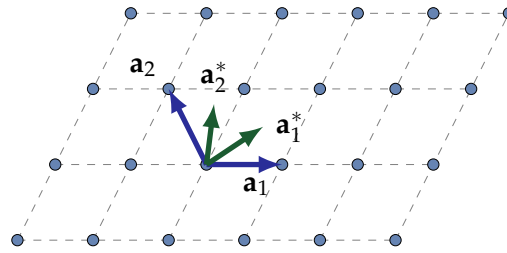
Figure 3.4: Real and reciprocal lattice vectors.

That is the set of diffraction vectors forms a lattice, i.e. a periodic array. Because the vectors magnitude is proportional to the reciprocal of the lattice spacing the vectors $\mathbf{a}_1^*$, $\mathbf{a}_2^*$ and $\mathbf{a}_3^*$ are called **reciprocal lattice vectors** and the set of all diffraction vectors are called the reciprocal lattice.

3.3 Physical Reality of Reciprocal Lattice Vectors

As the method by which the reciprocal lattice vectors are derived is very mathematical you might be tempted to think that they don't have a physical reality in the same way that lattice vectors do. This is wrong. In fact there is a one-to-one correspondence between reciprocal lattice vectors and planes of identical atoms within the crystal, with the reciprocal lattice vector being the normal to its corresponding plane.

The relationship between reciprocal lattice vectors and crystal planes means that if you rotate a crystal you are rotating its reciprocal lattice vectors in just the same way that you are rotating it's lattice vectors.

3.4 Measuring Reciprocal Lattice Vectors using a Single Crystal X-ray Diffractometer
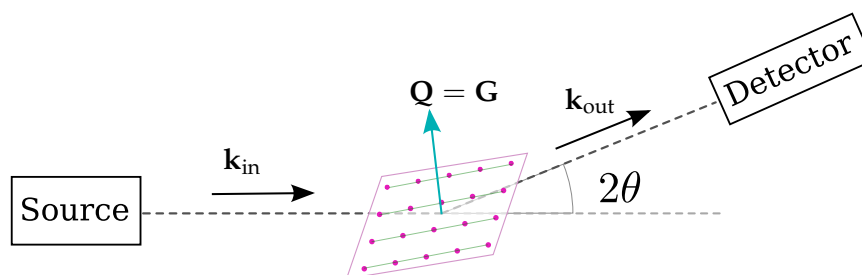
Figure 3.5: X-ray diffraction geometry showing the incident and reflected wave vectors ($\mathbf{k}_{in}$ and $\mathbf{k}_{out}$)

From the derivation set out above we know that in order to see diffraction of X-rays from a crystal the incoming and outgoing X-ray wavevectors have to satisfy the equation $\mathbf{Q} = \mathbf{k}_{out} - \mathbf{k}_{in} = h\mathbf{a}_1^* + k\mathbf{a}_2^* + l\mathbf{a}_3^*$, i.e. the difference in the X-ray wavevectors before and after scattering is equal to a reciprocal lattice vector. Thus X-ray diffraction allows us to measure

the reciprocal lattice vectors of a crystal and from these the lattice vectors. In order to make such a measurement we need to determine the incoming and outgoing X-ray wavevectors. This is done by defining the direction of the X-rays by using slits to collimate the X-ray beam going from the X-ray source to the sample and the X-ray beam from the sample going to the detector. Note that the detector and its slits are normally mounted on a rotation stage allowing the direction of $\mathbf{k}_{out}$ to be controlled. The magnitude of $\mathbf{k}_{in}$ is given by the energy of the X-rays from the source, which is normally chosen to be monochromatic. The magnitude of $\mathbf{k}_{in}$ is given by the fact that as the energy of the X-rays is much higher than the energy of the excitations of the crystal X-ray scatting can basically be treated as elastic i.e. $|\mathbf{k}_{out}| = |\mathbf{k}_{in}|$.

By controlling the angle, 2θ, between the incoming and outgoing beams we control the magnitude of the scattering vector $\mathbf{Q}$. In order to get diffraction from a specific reciprocal lattice vector we have to set this angle to a specific value. This angle is determined by the Bragg formula:

$$\frac{2\pi}{\lambda_{\text{X-rays}}} 2\sin(\theta) = |h\mathbf{a}_1^* + k\mathbf{a}_2^* + l\mathbf{a}_3^*| \tag{3.9}$$

However controlling 2θ is not enough to ensure we get diffraction in addition we have to rotate the crystal so that one of its reciprocal lattice vectors is in the direction of $\mathbf{Q}$.
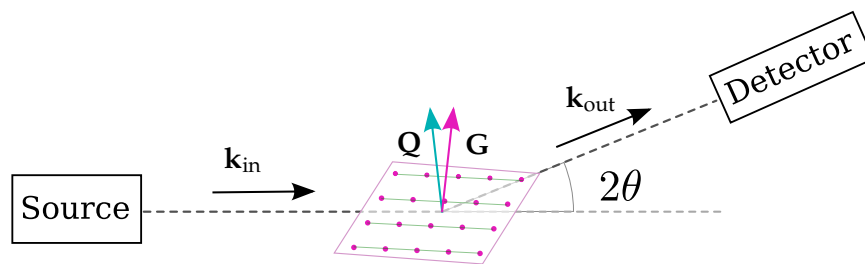
Figure 3.6: Crystal not rotated correctly therefore no diffraction observed.

For this reason in a single crystal x-ray diffractometer the sample is also mounted on a rotation stage which allows it to be rotated about 2 different axis by angles ω (omega) ϕ (phi) respectively. This is why single crystal diffractometers are called 3 axis diffractometers. Below is a picture of a diffractometer currently being used in Southampton.
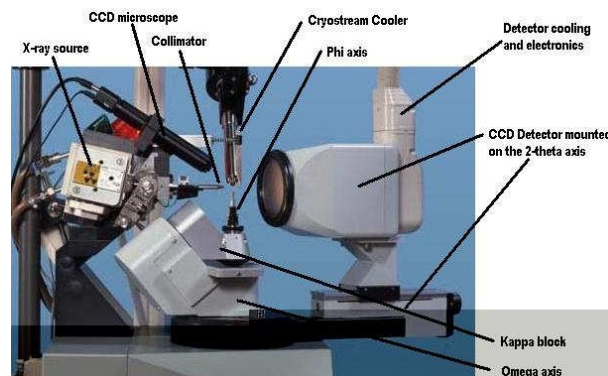
Figure 3.7: Single crystal X-ray diffractometer.

3.5   The Effect of the Unit Cell/Basis - The Structure Factor

So far we have only calculated the effect of the lattice on the diffraction pattern. What happens if we include the effect of the unit cell? As the lattice and unit cells terms are multiplied together it is clear that diffraction will only occur if the scattering vector is equal to a reciprocal lattice vector. This means that the effect of the unit cell can only be to alter the relative intensity of the scattering from different reciprocal lattice vectors. The effect of the unit cell occurs via the integral over a single unit cell which is called the structure factor:

$$S(\mathbf{Q}) = \iiint_{\text{UnitCell}} \rho_e(\mathbf{r})\, e^{-i\mathbf{Q}\cdot\mathbf{r}} dV \tag{3.10}$$

In order to try and understand what effect the atoms in the unit cell we will consider two simple models. In the first we will have the simplest type of unit cell, i.e. a single atom. We will treat the atom as being a point like object with all of its electrons at its nucleus. Thus the electron density in the unit cell is given by $\rho_e(\mathbf{r}) = N_e\delta(\mathbf{r})$ where N_e is the number electrons the atom has and $\delta(\mathbf{r})$ is the Dirac delta function, i.e. a big spike at the origin. The result of the integral in this case is given by:

$$\iiint N_e\delta(\mathbf{r})\, e^{-i\mathbf{Q}\cdot\mathbf{r}} dV = N_e \tag{3.11}$$

Thus for a single atom the X-ray intensity is the same for all of the reciprocal lattice vectors. Next we will investigate the effect of the unit cell consisting of two different atoms with one at the origin and the other at position $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$ (as described in the previous section: positions of atoms with the unit cell are normally defined in terms of the lattice vectors). In this case:

$$\rho(\mathbf{r}) = N_e^1\delta(\mathbf{r}) + N_e^2\delta\left(\mathbf{r} - \frac{1}{2}\mathbf{a}_1 - \frac{1}{2}\mathbf{a}_2 - \frac{1}{2}\mathbf{a}_3\right) \tag{3.12}$$

and:

$$S(\mathbf{Q}) = \iiint_{\text{UnitCell}} \rho_e(\mathbf{r})\, e^{-i\mathbf{Q}\cdot\mathbf{r}} dV = N_e^1 + N_e^2 e^{-i\mathbf{Q}\cdot(\frac{1}{2}\mathbf{a}_1+\frac{1}{2}\mathbf{a}_2+\frac{1}{2}\mathbf{a}_3)} \tag{3.13}$$

as diffraction only occurs when $\mathbf{Q} = \mathbf{G} = h\mathbf{a}_1^* + k\mathbf{a}_2^* + l\mathbf{a}_3^*$ this expression can be rewritten:

$$S(\mathbf{Q}) = \iiint_{\text{UnitCell}} \rho_e(\mathbf{r})\, e^{-i\mathbf{Q}\cdot\mathbf{r}} dV = N_e^1 + N_e^2 e^{-i\pi(h+k+l)} \tag{3.14}$$

If $h+k+l$ is even then the exponential equals one and the intensity will equal $(N_e^1 + N_e^2)^2$. If $h+k+l$ is odd then the exponential equals minus one and the intensity will equal $(N_e^1 - N_e^2)^2$. Thus if we were to measure the intensity of the different diffraction peaks we could determine the relative number of electrons on each atom (as the number of electrons will be integers this is often enough to allow the absolute numbers to be determined).

Of course in reality atoms are not points but instead have a finite size, however by measuring the intensity of diffraction for different reciprocal lattice vectors it is possible to determine a lot about the electron density within the unit cell.
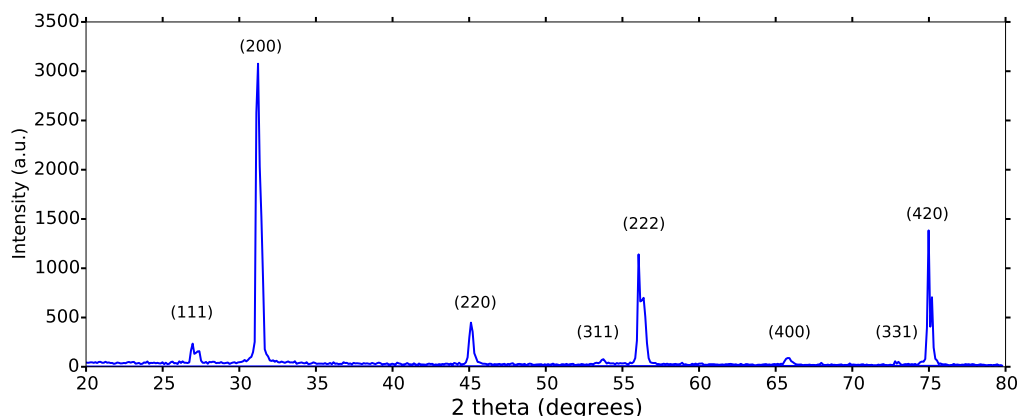
Figure 3.8: Diffraction pattern from a NaCl crystal. The different peak heights are due to the fact that Na^+ and Cl^- don't have the same number of electrons.

3.6 Powder Diffraction Technique

As most of the possible values of the 3 angles which can be varied in a 3 axis X-ray diffractometer will give absolutely no signal and until we've done a diffraction experiment we probably won't know which direction the lattice, and therefore reciprocal lattice, vectors are pointing aligning a 3 axis experiment can be very time consuming. To get round this problem a more common way to measure diffraction patterns is to powder the sample up. A powdered sample is effectively a vast number of small crystals orientated at random. This means that for any orientation of the sample at least some of the small crystals will be orientated so that their reciprocal lattice vectors are pointing in the direction of the scattering vector, **Q**, and in this case to get diffraction it is only necessary to satisfy Bragg's law in order to get diffraction.

3.7 Properties of Reciprocal Lattice Vectors

Although we have calculated the reciprocal lattice vectors by considering diffraction of X-rays they come about due to the periodicity of the lattice and therefore they are extremely important in many different aspects of discussing waves in crystals. At this stage in the course we will look at one of them however we will return to this topic several times later on.
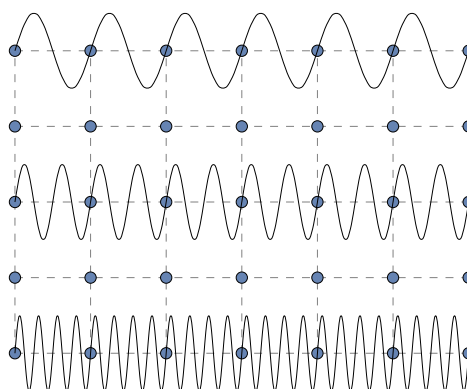
Figure 3.9: Waves in a crystal.

3.7.1 Periodic Functions

The defining property of a crystal is translational symmetry. That is if you move a crystal by a lattice vector it is indistinguishable from before you moved it. An obvious example is that the electrostatic potential produced by the nuclei of the atoms making up the sample has full translational symmetry i.e. it is periodic with the period of the lattice. In addition as we will go on to show later this implies that in a crystal the charge density of the eigenstates of electrons have to have the same symmetry.

Because of the way we defined the reciprocal lattice vectors all waves with a wavevector which are a reciprocal lattice vector will have the same phase at two equivalent points in any unit cell (see fig. 3.9). In fact all waves which have this property have wavevectors which are reciprocal lattice vectors. Fourier theory tells us that any function which is periodic can be represented by a sum of all the sinusoidal waves which have the property that their period is such that they have the same phase at two equivalent points in any two repeat units, i.e. unit cells.

As many of the properties of a crystal are periodic in the lattice and all such periodic functions can be represented as a sum of sinusoidal waves:

$$m(\mathbf{r}) = \sum_{\mathbf{k}} c_{\mathbf{k}} \, e^{i\mathbf{k}\cdot\mathbf{r}} \tag{3.15}$$

Reciprocal lattice vectors are extremely important to nearly all aspects of the properties of materials.

3.7.2 Prelude to Lattice Vibrations

Reciprocal lattice vectors are also important when we discuss sounds waves. Normally when we discuss sound in solids we are thinking about audible frequency waves with a long wavelength and therefore small wavevector. In this case it is acceptable to treat the solid as if it was continuum and ignore the fact it is made up of discrete atoms. However actually solids are made up of atoms and therefore at the smallest scales the motion caused by the sound wave has to be described in terms of the motion of individual atoms. In this case the wavevector of the sound $\mathbf{k}$ tells us the phase difference between the motion of the atoms in different unit cells. i.e. if the phase of oscillation in one unit cell is defined as zero then the phase in another unit cell which is a position $\mathbf{R}_n$ relative to the first, where $\mathbf{R}_n$ is therefore a lattice vector, is given by $\mathbf{k} \cdot \mathbf{R}_n$. Now if we define a new sound wave whose wavevector is given by $\mathbf{k} + \mathbf{q}_R$ where $\mathbf{q}_R$ is a reciprocal lattice vector then the phase in the second unit cell is given by $\mathbf{k} \cdot \mathbf{R}_n + \mathbf{q}_R \cdot \mathbf{R}_n = \mathbf{k} \cdot \mathbf{R}_n + 2\pi \times$ integer $= \mathbf{k} \cdot \mathbf{R}_n$. As this is truly independent of $\mathbf{R}_n$ the first and second sound waves are indistinguishable and in fact there is only one soundwave which is described by both wavevectors. As this is truly independent of which reciprocal lattice vector is chosen there are an infinite set of wavevectors which could be used to describe any particular soundwave. In order to bring some sanity to this situation there is a convention that we only use the smallest possible wavevector out of the set.

Like all other vector quantities, wavevectors can be plotted in a space and therefore the set of all possible wavevectors which represent soundwaves can be plotted as a region in this space. This space is called the 1st Brillouin zone. The 1$^{\text{st}}$ Brillouin zone can be easily calculated as it is the Wigner-Seitz primitive unit cell of the reciprocal lattice (see previous section).

3.7.3 Electron Wavefunctions

As we will prove later on in the course the wavefunction of an electron in a solid can be written in the form:

$$\psi(\mathbf{r}) = u(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}} \tag{3.16}$$

where $u(\mathbf{r} + \mathbf{R}) = u(\mathbf{r})$ for all lattice vectors $\mathbf{R}$, i.e. that $u(\mathbf{r})$ is the same at every lattice point. If we again add a reciprocal lattice vector to the wavevector then:

$$\psi(\mathbf{r}) = u(\mathbf{r})e^{i(\mathbf{k}+\mathbf{q})\cdot\mathbf{r}} = \left(u(\mathbf{r})e^{i\mathbf{q}\cdot\mathbf{r}}\right)e^{i\mathbf{k}\cdot\mathbf{r}} = u_{\text{new}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}} \tag{3.17}$$

i.e. we can represent all possible electron wavefunctions only using wavevectors within the 1st Brillouin zone.

3.8 Electron and Neutron Diffraction

As with all quantum matter, electrons and neutrons exhibit wave-like phenomena and can scatter from a crystal when their wavelength is comparable to the inter-atomic spacing. A different interaction mechanism is however responsible for diffraction effects in each case. Electrons are scattered by the potential produced by the nucleus and the outer electrons in contrast to X-rays which are scattered primarily by core electrons. Electrons are scattered by a factor of 10^4 times more strongly than X-rays and the resulting diffraction pattern is readily observed using fluorescent imaging or photographic film. Low energy (10 - 500 eV) electron beams are rapidly attenuated as they enter the solid and are therefore amenable to the study of surfaces and two-dimensional structures. The analysis that was developed for X-ray diffraction above is only applicable if we assume that the wave scattered by each atom is sufficiently weak so that rescattering effects can be ignored. In principle the wavelength for electrons is continuously variable and can be tuned by accelerating electrons through a potential. To obtain a wavelength of 1 Å, and electron needs an energy of $\sim$150 eV.

Neutrons are scattered by the nucleus of the sample due to the strong nuclear force. The neutron is momentarily captured by the nucleus and then re-emitted. If the nucleus has a finite spin, then the compound nucleus will either increase or decrease its spin by $\frac{1}{2}$. As a result, scattering from the material will contain a coherent and an incoherent component. In some cases, the incoherent component is dominant, as is the case for Vanadium. The strength of the scattering does not increase monotonically with atomic number as is generally the case for X-rays. Neutrons are therefore more useful for diffraction experiments involving light elements (such as hydrogen) and for comparison of elements with comparable atomic number. As the nucleus is much smaller than the neutron wavelength, it acts as a points scatterer and there is no angular dependence on the strength of scattering. Neutrons also possess a net magnetic moment and can therefore interact with magnetic atoms. This is useful to study magnetic structure in a variety of compounds consisting of transition and rare earth metals.

X-rays, electrons and neutrons can be used to investigate the dispersion relationship of excitations such as phonons in solids. In order to do so we must measure the frequency (energy) change of the particle as it interacts with the material. As mentioned above, the wavelength

must be comparable to the atomic spacing. For a wavelength of 2 Å, we obtain the following energies:

X-rays	Electrons	Neutrons
$E = h\nu = \frac{hc}{\lambda}$	$E = \frac{\hbar^2 k^2}{2m_e}$	$E = \frac{\hbar^2 k^2}{2m_n}$
$\simeq 5 \times 10^4$ eV	$\simeq 40$ eV	$\simeq 0.02$ eV

Phonons typically have energies in the range of 0.01 - 0.1 eV. Comparing this with the energies tabulated above, we can see that a significantly higher energy resolution is needed to measure energy changes with X-ray photons as opposed to neutrons. If a photon of lower energy is used, its wavelength is no longer comparable to that of the interatomic spacing and therefore diffraction effects are not observed. Neutrons might appear to be a superior tool to X-ray sources. This however is not always the case as compact and bright X-ray sources are widely available. Neutrons are obtained from nuclear fission reactors or spallation sources which are large and very expensive. In addition, the neutron intensity from these sources is relatively low and the range of energies is large. In contrast, intense monochromatic X-rays are readily obtained as characteristic emission lines from an X-ray tube and are normally used where possible.

Chapter 4

Crystal Dynamics

The physical description of a crystal presented in previous chapters is that of a collection of atoms that remain stationary when condensed into a regular arrangement, or crystal structure. We know that this is not completely representative of real crystals as the uncertainly principle does not allow the position and momentum of a particle to be determination simultaneously. This means that the atoms in a crystal will vibrate even at a temperature of absolute zero. Physical material properties related to the lattice dynamics include the velocity of sound waves in the material, the specific heat capacities, thermal expansion, thermal conductivity (insulators) and mechanical hardness.

Amongst the most elementary descriptions of atomic motions in a lattice is the Einstein model. In this model, each atoms is permitted to vibrate independently about an equilibrium point due to the potential well created by neighbouring atoms. This model is useful at high temperatures, where the assumption that the atoms are able to vibrate independently can be justified. The model fails however to accurately describe thermal properties at low temperature and predicts a more rapid decrease in heat capacity. We can however see that if two or more neighbouring atoms move in unison, the forces that would tend to restore the atoms back to equilibrium would be greatly reduced. Moreover, there is a tendency for the motion of adjacent atoms to be correlated. Such effects are not accounted for by a linear harmonic oscillator model. Further refinements were made when Debye introduced the concept of a frequency spectrum of the elastic continuum within which the material is permitted only a fixed number of vibrational modes. This modification to the Einstein model agrees more closely with experimental data at low temperatures.

Both the Einstein and Debye models can only approximate the dynamics of a lattice as they are both based on a continuous solid. An atomistic theory of lattice dynamics provides the most elementary means to determine the frequency spectrum of collective oscillations of a crystal lattice and is also able to provide a connection between the frequency and wavenumber in specific directions (i.e. the dispersion relationship). In the following we shall consider a simple one-dimensional lattice models of a crystal as a means to investigate the characteristic behaviour of real crystals.

4.1 Lattice Vibrations

To begin our discussion of lattice vibrations, we consider that atoms displaced from their equilibrium positions will induce a change in the distribution of electrons and a new ground state energy for that atomic configuration. If the atoms are restored to the initial positions the initial ground state energy is recovered. The total energy of the system is then a function of the coordinates of all atoms and can be considered a potential for the atomic motion.

4.1.1 One Atom per Unit Cell

We can deduce this potential for a simple one-dimensional crystal consisting of a chain of N identical atoms each of which is separated by a lattice constant a. We assume that atoms are only free to move in the direction parallel to the axis of the chain. We also assume that the forces between atoms are short range so that only nearest neighbour interactions are important. A Taylor series expansion about $r = a$ of the potential $W(r)$ for a single atom is given by:

$$W(r) = W(a) + \frac{(r-a)^2}{2}\frac{\mathrm{d}^2 W(a)}{\mathrm{d}r^2} + \dots .$$

(4.1)

Higher order terms are ignored in what is known as the *harmonic approximation*. Terms linear in $(r - a)$ disappear as the expansion is about the equilibrium position. The force on each atom due to a displacement $u = r - a$ from the equilibrium position is then:

$$
\begin{aligned}
F &= -\frac{\mathrm{d}W}{\mathrm{d}u} = -cu \\
c &= \frac{\mathrm{d}^2 W(a)}{\mathrm{d}r^2}.
\end{aligned}
$$

(4.2)

Displacement of an atom with index $s + p$ creates a force on atom s in the direction of that displacement. The total force that the atom s experiences due to its immediate neighbour at $s + p$ is the then given by:

$$F_s = c_p(u_{s+p} - u_s).$$

(4.3)

If each atom has a mass of M, the equation of motion for atom s is obtained by summing the contribution of all neighbouring atoms:

$$M\frac{\mathrm{d}^2 u_s}{\mathrm{d}t^2} = \sum_p c_p(u_{s+p} - u_s).$$

(4.4)

If we assume that all atoms oscillate with the same amplitude u_0, we can represent the displacement at s as a travelling wave given by:

$$
\begin{aligned}
u_s &= u_0 e^{i(ksa - \omega t)} \\
u_{s+p} &= u_0 e^{i(k(s+p)a - \omega t)}.
\end{aligned}
$$

(4.5)

Substituting into equation 4.4 and noting the symmetry of the lattice ($c_p = c_{-p}$) yields:

$$-M\omega^2 = \sum_{p>0} c_p(e^{i(kpa)} + e^{-i(kpa)} - 2).$$

(4.6)

The above can be written as:

$$\omega^2 = \frac{2}{M} \sum_{p>0} c_p (1 - \cos kpa) = \frac{4}{M} \sum_{p>0} c_p \sin^2 \frac{1}{2} kpa. \tag{4.7}$$

For nearest neighbour interaction only ($p = 1$) we have:

$$\omega^2 = \frac{4c_1}{M} \sin^2 \frac{1}{2} ka. \tag{4.8}$$

This is the dispersion relationship lattice vibrations in a linear chain of atoms. There is a maximum cut-off frequency of propagation given by $2(c_1/M)^{1/2}$, as postulated in the Debye model. Notice also that s has cancelled out and equation 4.8 is therefore valid for all atoms in the lattice. As oscillations in each atom are coupled (Eq. 4.3) an atom that starts to vibrate will not continue to do so at a constant amplitude but will instead dissipate energy to neighbouring atoms. The allowed frequencies are found to vary periodically with k on an interval of $2\pi/a$. We should also expect the number of propagating modes to be given by:

$$k = \frac{2\pi}{\lambda} = \frac{2\pi p}{Na}. \tag{4.9}$$

There are therefore N possible values of k in the range $2\pi/a$, say $-\pi/a < k \leq \pi/a$. Waves outside this range ($\lambda < 2a$) are unphysical and contain no new information beyond that contained within the first Brillouin zone. Waves with $k = \pm\pi/a$ ($\lambda = 2a$) will undergo Bragg reflection (and result in standing waves) with a Bragg angle of $\pi/2$.[1]

In addition to longitudinal lattice vibrations, the linear lattice can also support two mutually transverse displacements which leads to an independent dispersion relationship for each that describes transverse wave propagation along the lattice. In a linear lattice, these modes are degenerate.

4.1.2 Two Atoms per Unit Cell

In the previous discussion of a linear lattice, all atoms are assumed to be equivalent and there is therefore only one atom per unit cell. In this case, we expect that within the local vicinity of a few atoms, the motion of the atoms will be correlated causing atoms to move in the same direction. This type of motion is characteristic of sound waves and the corresponding vibrational modes are known as the acoustic modes or branches. A somewhat different situation arises when we consider a linear lattice consisting of two non-equivalent atoms. In this case, two dissimilar atoms occupy non-equivalent positions within the unit cell of the lattice. This is the simplest model of an ionic crystal and serves as a means to deduced the important qualitative features. Following the same reasoning above, if a is the lattice spacing and $a/2$ is the nearest neighbour separation, the equations of motion are written down as:

$$
\begin{aligned}
M_1 \frac{d^2 u_s}{dt^2} &= \sum_p c_p (u_{s+p+1/2} - u_s) \\
M_2 \frac{d^2 u_{s+1/2}}{dt^2} &= \sum_p c_p (u_{s+p} - u_{s+1/2}).
\end{aligned}
\tag{4.10}
$$

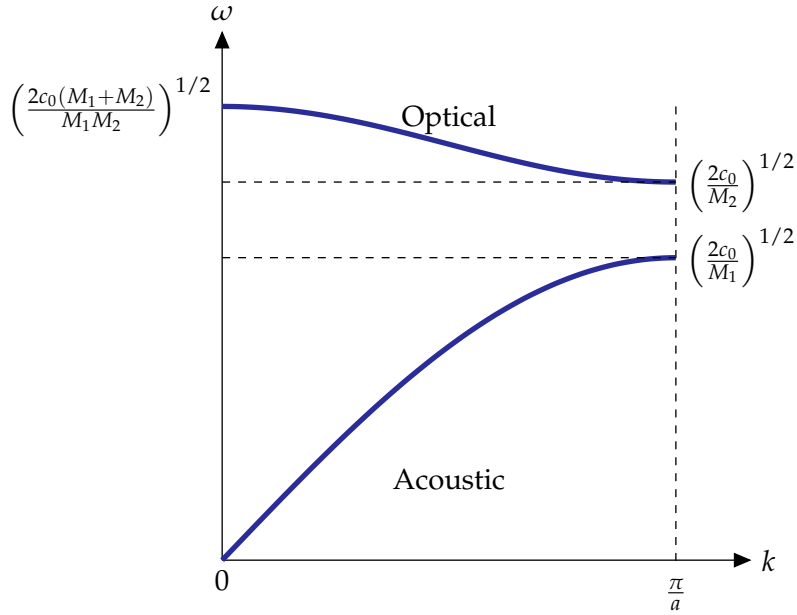[1] A Bragg angle of $\pi/2$ is equivalent to a reflection along the direction of propagation.

Figure 4.1: Optical and acoustic mode branches of the lattice vibration dispersion curve.

If the relative amplitude of oscillation is given by α, the displacement for atom $s + 1/2$ is given by:

$$u_{s+p+1/2} = \alpha u_0 \, e^{i(k(s+p+1/2)a - \omega t)}. \tag{4.11}$$

which yields the following set of coupled equations ($c_p = c_{-p}$, for $p \in [0, 1, 2, ...]$):

$$
\begin{aligned}
-M_1 \omega^2 &= \sum_{p=0}^{N} c_p (\alpha \, e^{i(k(p+1/2)a)} + \alpha \, e^{i(k(p-1/2)a)} - 2) \\
-\alpha M_2 \omega^2 \, e^{i(k(1/2)a)} &= \sum_{p=0}^{N} c_p (e^{i(kpa)} + e^{-i(kpa)} - 2\alpha \, e^{i(k(1/2)a)}).
\end{aligned}
\tag{4.12}
$$

For nearest neighbour interaction we can simplify the above to give:

$$
\begin{aligned}
-M_1 \omega^2 &= 2c_0 \left\{ \alpha \cos\left(\frac{ka}{2}\right) - 1 \right\} \\
-\alpha M_2 \omega^2 &= 2c_0 \left\{ \cos\left(\frac{ka}{2}\right) - \alpha \right\}.
\end{aligned}
\tag{4.13}
$$

Eliminating α and solving for ω yields:

$$\omega_\pm^2(k) = c_0 \left(\frac{M_1 + M_2}{M_1 M_2}\right) \pm \left\{ c_0^2 \left(\frac{M_1 + M_2}{M_1 M_2}\right)^2 - \frac{4c_0^2}{M_1 M_2} \sin^2 \frac{ka}{2} \right\}^{1/2} \tag{4.14}$$

There are two solutions for $\omega_\pm^2(k)$ which results in two branches in the dispersion relationship. The lower branch can be associated with acoustic cooperative modes ($\omega_-^2(k)$), which are equivalent to those discussed previously. The upper branch is associated with the movement of atoms in opposing directions and is referred to as the optical branch ($\omega_+^2(k)$). This is because such modes cause an electrical polarisation in ionic crystals such as CsCl that can couple to light and cause strong optical absorption. Optical modes reside at higher frequencies and have a dispersion curve that is weaker than the acoustic branch. The two groups of modes
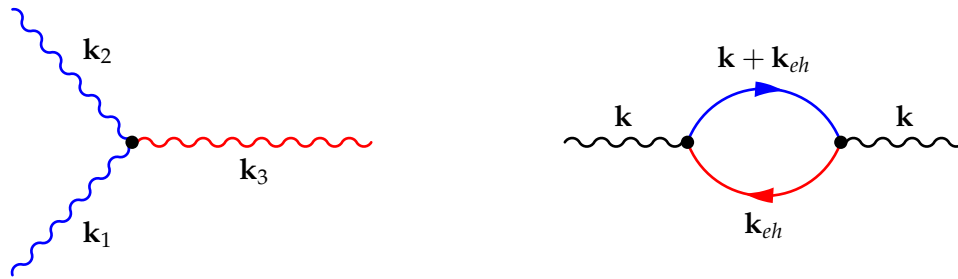
Figure 4.2: Left: Interaction of phonons conserving energy and wavevector. Right: Phonon self energy caused by creation of a virtual quasi-electron-hole pair. This process leads to softening of frequency modes and can be responsible for structural phase transitions.

are separated by intervals in frequency (or equivalently, energy) that prohibit collective lattice vibrations.

4.1.3   Phonons

It is often convenient to introduce the concept of the quanta of a lattice vibration known as the phonon. Phonons are quasi-particles that can be viewed as having wave-like and particle-like properties.[2] Phonons are not localised particles. The is because the plane wave used to describe them extends throughout the crystal and has a well defined momentum $\hbar\mathbf{k}$. Just as with photons, however, we are able to construct a wavepacket that is localised over a few unit cells by combining modes within a reasonably narrow range. This wavepacket will therefore represent a localised phonon moving with group velocity $\mathrm{d}\omega/\mathrm{d}\mathbf{k}$.

It is often convenient to associate the quantity $\hbar\mathbf{k}$ with the momentum of a phonon. As vibrational modes in a crystal result from atomic motion about equilibrium positions and there is no rigid motion of the crystal, the average displacement is zero and therefore the quantity $\hbar\mathbf{k}$ does not represent conventional linear momentum. Vibrational modes however do interact and are subject to both energy conservation and also conservation of the quantity $\hbar\mathbf{k}$ - known as *crystal momentum*. Interacting phonons conserve crystal momentum according to the following laws:

$$\hbar\omega_1 + \hbar\omega_2 \; = \hbar\omega_3$$

(4.15)

$$\hbar\mathbf{k}_1 + \hbar\mathbf{k}_2 \; = \hbar\mathbf{k}_3 + \hbar\mathbf{G}$$

where $\mathbf{G}$ is a reciprocal lattice vector. We shall revisit this concept in later chapters in the context of electron quasi-particles.

4.2   Anharmonic Interactions (Not Examinable)

In real crystals, there is a departure from the interaction potential describe in equation 4.1. This is understood if we consider that a crystal will resist compression far more greatly than it will resist expansion into a larger volume. This asymmetry is not found in harmonic approximation of equation 4.1. Other properties that are not predicted by the harmonic ap-

---

[2]Quantisation of phonons is formally achieved using the method of Second Quantisation.
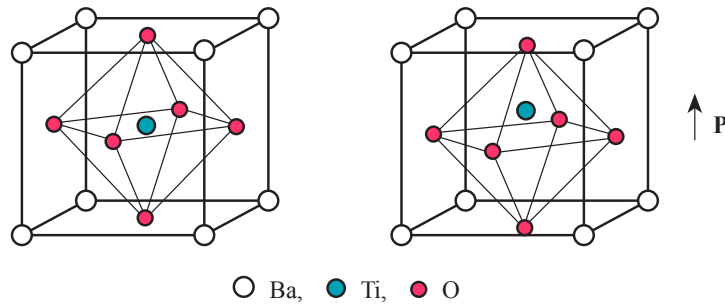
Figure 4.3: Conventional unit cell of barium titanate (BaTiO$_3$) showing a structural phase transition. The electric polarization is brought about by the upward displacement of the central Ti^{4+} atom (ferroelectricity). The polarization **P** is a natural order parameter.

proximation are known as anharmonic effects and the term in equation 4.1 required for their correct description are known as anharmonic terms. Anharmonicity causes renomalisation of the mode frequencies (also known as "dressing") and decay into other modes (Fig. 4.2). An important effect that necessitates anharmonic terms in the crystal potential is thermal expansion. The harmonic approximation to the interaction potential fixes the mean separation of the atoms while their vibration amplitudes can increase with temperature. Inclusion of higher order term in equation 4.1 results in the coupling of modes and corresponds to scattering of phonons which limits the thermal conductivity associated with phonon transport. Without this correction, phonons do not interact.

Anharmonic interactions are also important for the study of structural phase transitions in crystalline materials. Phonons can be "dressed" by the creation of virtual quasi-electron-hole pairs which results in the softening of the mode frequencies (Fig. 4.2). A soft mode occurs when the corresponding frequency of a mode is vanishingly small at a particular temperature. This causes atomic displacements associated with that mode to become time-independent and a rigid displacement of atoms within the unit cell occurs. Mode softening therefore plays a central role in the dynamic structural reordering of crystals.

Mode coupling also occurs when different types of collective excitations (such as phonons) interact. Examples include optical phonon-photon (phonon polaritons) coupled excitations, optical phonon-plasmon (polarion) coupled excitations and exciton-photon (exciton polariton) coupled excitations. In each case, where the two modes intersect in the dispersion curve, they become mixed in approximately equal amounts and each of the collective excitations are no longer considered distinct.

4.3   Structural Phase Transitions (Not Examinable)

The concept of a phase transition is usually encountered in the context of a thermodynamic system that undergoes a transitions from solid to liquid and then to a gas as certain properties of the system are altered. Phase transitions can be driven by a wide range of parameters including pressure, temperature, electric field, magnetic field, impurities and chemical composition. In this general category of a thermodynamic system, the microscopic constituents can be considered as non-interacting and as a result the thermodynamic properties of the system are directly inferred from a knowledge of the energy levels of the individual constituents.

Table 4.1: Order parameters for different phase transitions.

Phase transition	Symmetry broken	Order Parameter
Gas - Liquid	Reflection	$\rho_g - \rho_l$
Liquid - Solid	Reflection	Pair correlation function
Solid - Solid	Lattice symmetry	Fourier transform of electron density
Ferroelectricity	Lattice symmetry	Polarisation P
Liquid crystal	$SO(3)/Z_2$ rotation	Orientation
Superconductivity	U_1 gauge	Energy gap
Superfluidity	U_1 gauge	Wavefunction

Examples of this include specific heats of gases, chemical reactions and condensation of an ideal Bose gas. The remaining category of thermodynamic system is characterised by analytic discontinuities in the thermodynamic functions of the system. Example phenomena include condensation of gases, melting of solids, mixing of liquids, ferromagnetism and superfluidity in liquid He. Phase transitions in this case exhibit a tendency towards strong and cooperative interaction between the constituents at a critical temperature (T_c) or some other transition parameter.

A phase transition is invariably accompanied by a change in entropy of the system. With each phase of the system there is an associated Gibbs potential $G(p, T)$ that indicates the most stable phase at the minimal possible value for a given condition. A phase transition can be further classified as either *discontinuous* or *continuous* across the transition region of the phase diagram. Discontinuous transitions are considered as first-order and are accompanied by a release of latent heat during the transitions. All other thermodynamic properties are also discontinuous. Continuous transitions are considered second-order and have continuous thermodynamic properties, although their first derivatives are discontinuous. An example is the onset of magnetic ordering below the Curie temperature.[3]

Phase transitions can also occur in crystalline solids. In this case, the crystal remains in the solid phase after the transition has occurred. Crystals can undergo both first-order and second-order phase transitions. First-order structural phase transitions are associated with a complete reordering of the lattice structure. The structural transition of Fe from bcc to fcc lattice structure is a notable example. Also, the structural phase transition of VO_2 from monoclinic to tetrahedral is of technological importance. Second-order structural phase transitions are not associated with a change in crystal structure but rather a distortion from a regular arrangement. This occurs when the internal symmetry of the unit cell is altered in a continuous way. A characteristic feature of this is the bifurcation of the Gibbs potential into two branches below the ordering temperature corresponding to the ordered and disordered states. An example of this is the perovskite crystal structure of barium titanate. This material undergoes a continuous second-order structural phase transition at 405K. This is signified by an upward displacement of the central Ti^{4+} ion parallel to one of the cubic axes (Fig. 4.3).

Mathematical problems and concepts associated with the study of cooperative phenomena in first-order phase transitions are formidable at best and extend far beyond the scope of this course. The study of continuous second-order phase transitions is rather more accessible and lends itself to a phenomenological study. We will turn our attention to this in the next section.

---

[3]See chapter 8

4.4   Landau Theory and the Order Parameter (Not Examinable)

Landau theory was developed as a general means to describe second-order phase transitions. In his theory, Landau postulated that the thermodynamic free energy should be analytic, and that it should obey the symmetry of the Hamiltonian. His insight was to expand the thermodynamic free energy as a power series of an 'order parameter' η. A phase transition can then occur when one or more coefficients of the power series changes sign. The order parameter η is a measurable quantity which is zero above T_c (critical temperature) and non-zero below T_c. The emergence of an order parameter is accompanied by certain symmetry breaking. Different systems exhibit differing phase transitions and as a result have different kinds of order parameters. The order parameter can be a scalar, a complex number, a vector, a tensor, or a more abstract quantity. Examples of order parameters for different phase transitions are listed in table 4.1. The order parameter is often determined by experimental means or from observational insight. Once known, it is possible to replace in the Hamiltonian the microscopic operator with its mean-field average and neglect any fluctuations.

Landau further postulated that the Gibbs potential can be expanded into an infinite power series of η, which is expressed as

$$G(\eta, T) = \sum_i^n \frac{1}{2i} c_{2i}\eta^{2i} = c_0 + \frac{1}{2}c_2\eta^2 + \frac{1}{4}c_4\eta^4 + \frac{1}{6}c_6\eta^6 + ... + \frac{1}{2n}c_{2n}\eta^{2n} \tag{4.16}$$

where the coefficients c_{2i} are normally smooth functions of temperature. Odd power terms are not included in the expansion as ordered states in an isotropic system are invariant by inversion or reflection of η in a mirror plane. As a result $G(\eta) = G(-\eta)$. The series is usually truncated upto the fourth power without loss of accuracy. The quartic term in the potential is generally related to microscopic correlations within the system. The condition for thermal equilibrium of the system also yields the order parameter and is given by:

$$0 = \frac{\partial G(\eta, T)}{\partial \eta} = c_2\eta + c_4\eta^3 \tag{4.17}$$

which yields:

$$\eta = \begin{cases} 0, & T > T_c, \quad c_2 > 0; c_4 = 0 \\[2mm] \pm\left(-\dfrac{c_2}{c_4}\right)^{1/2}, & T < T_c, \quad c_2 < 0; c_4 > 0 \end{cases} \tag{4.18}$$

The coefficient c_2 therefore changes to a negative value below T_c from positive value above T_c. This results in the equilibrium shifting to a non-zero value of η. Near the transitions region, the coefficient c_2 can be expanded around the critical temperature to yield:

$$c_2 = c_2'(T - T_c) \tag{4.19}$$

which also satisfies the requirements of Eq. 4.18. We can then determine properties of the system such as the entropy $S = -(\partial G/\partial T)_p$ or the heat capacity $C_p = -T(\partial S/\partial T)_p$.

Chapter 5

Free Electron Model - Jellium

Our first goal will be to try and understand how electrons move through solids. Even if we assume freely mobile electrons in the solid, we can predict many specific properties. We will start from what you already know from Quantum Mechanics and from the Quantum Physics of Matter course and work with that to see what it implies about the electronic properties of materials, and in particular, metals. The model we will use is incredibly simple, and yet it is very powerful and lets us make a first step in understanding how complex behaviours come about.[1]

5.1 Free Electron States

5.1.1 Jellium Potential

So our first, and most basic model will be a huge number of electrons, free to move inside a box. The box is the edges of our solid, because we know that electrons don't fall out of materials very easily (there is a large potential barrier). This must be a wild approximation since you know that solids are actually made of atoms. Rather amazingly, we can get a long way with this approximation. We will call our basic solid 'jellium', since the electrons can wobble around.
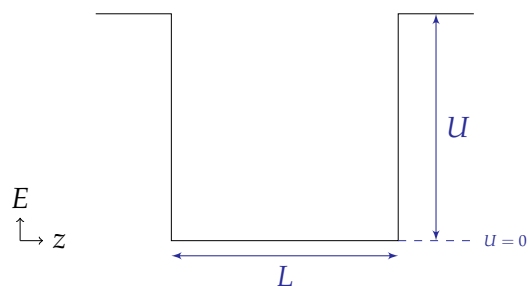
Figure 5.1: Constant potential solid in one-dimension. U is the potential in which all the electrons move and E is the kinetic energy.

---

[1]For additional material see Myers Chapters 6 and 7.

5.1.2 Solving Free Electron Wavefunctions

You can solve this problem using basic Quantum Mechanics: First you need the Schrödinger equation for electrons in free electron gas.

$$-\frac{\hbar}{2m}\nabla^2\psi + U\psi = E\psi \tag{5.1}$$

We have to assume that each electron doesn't need to know exactly where every other electron is before we can calculate its motion. There are so many electrons that each electron only feels an average from the others, and we include this in the total potential energy U.

This is a particle in a box problem, which you know how to solve. First you know that the momentum of the particle is a constant of the motion, i.e. the electrons are described by plane waves. This means they just bounce backwards and forwards against each side of the box, like the waves on a string. The solution is then a plane wave:

$$\psi = \frac{1}{\sqrt{V}}e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)} = \frac{1}{\sqrt{V}}e^{i(k_x x + k_y y + k_z z)}e^{-i\omega t} \tag{5.2}$$

The pre-factor ($V = L^3$, the volume of the box) is just a normalisation constant to make sure that the probability of finding the particle inside the box is unity. Given the wavefunction is $\psi = Ae^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}$, then we know the electrons are all in the sample, so

$$\int_{-\infty}^{\infty}\psi^*\psi dV = A^2 V = 1 \implies A = \frac{1}{\sqrt{V}} \tag{5.3}$$

The probability of electrons being in a certain volume is $\int \psi^*\psi dV = dV/V$. This implies that the probability for finding the electron just depends on the size of the volume we are considering compared to the total volume containing the electrons. So the electrons are uniformly distributed which is what the idea of plane waves would lead us to expect.
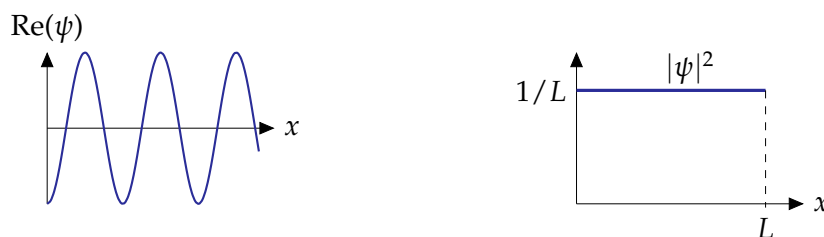
Figure 5.2: Plane wave states have equal probability density throughout space (1D in this example).

You should also remember that the energy of these plane wave solutions is just given by their kinetic energy (the bottom of the box is at an arbitrary potential) and equals:

$$E = \frac{\hbar^2 k^2}{2m} \tag{5.4}$$

This result is not particularly surprising since we have that the energy $\sim k^2$, which is just the kinetic energy of a free particle (momentum $\hbar k = mv$ and $E = mv^2/2$).

Figure 5.3: Dispersion relation of Jellium with constant energy contours projected onto the k_x-k_y plane in k-space.

We have drawn here (Fig. 5.3) the energy of the electrons as a function of their momentum, a dependence called their '**dispersion relation**'. This concept is pervasive throughout all physics because the dispersion relation dependence gives you a strong indication of how the waves or particles that it describes will behave in different conditions. For instanc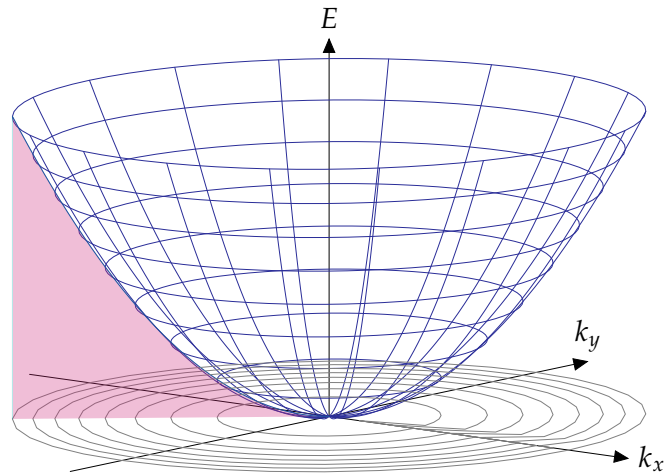e, if the dispersion relation is quadratic (as here) it gives us the feeling that the electrons are behaving roughly like billiard balls which are well described by their kinetic energy. In other cases you will see a dispersion relation which is extremely deformed from this (later in the course) in which case you know there are other interactions having an effect, and this will result in very different particle properties. Dispersion relations are also helpful because they encode both energy and momentum on them, and thus encapsulate the major conservation laws of physics!

The dispersion relation is hard to draw because there are three k axes, so we draw various cuts of it. For instance, Fig. 5.3 shows contours of constant energy of equal energy spacing, on the $k_x - k_y$ plane. Because the energy increases parabolically, the contours get closer together as you go further out in k.

As you saw last year, one useful thing you can extract from the dispersion relation is the **group velocity**. This is the speed at which a wavepacket centered at a particular k will move, $v_g = \frac{1}{\hbar} \frac{dE}{dk}$.

5.1.3   Boundary Conditions

The quantised part of the problem only really comes in when we apply the boundary conditions. Because the box has walls, only certain allowed values of k fit in the box. The walls cannot change the energy of the wave, but they can change its direction. Hence a wave e^{ikx} is converted by one wall into a wave e^{-ikx}, and vice-versa.

Hence for a 1-D case the wavefunction is a mixture of the forward and backward going waves:

$$\psi = \left(Ae^{ikx} + Be^{-ikx} \right) e^{-i\omega t} \tag{5.5}$$

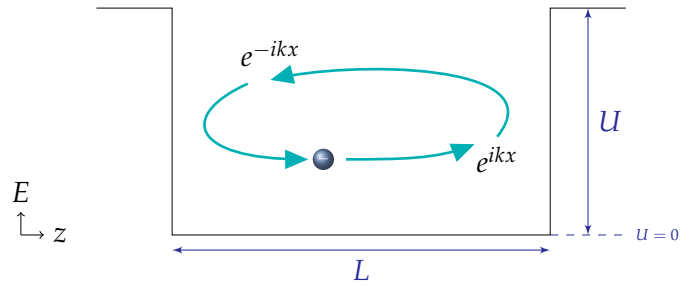When you put the boundary conditions for the full 3-D case:

Figure 5.4: Elastic collisions of electrons at the edges of the box.

$$\psi = 0 \text{ at } x = 0, y = 0, z = 0$$
$$\psi = 0 \text{ at } x = L, y = L, z = L$$

into (Eq. 5.5), the first says that $B = -A$, while the second implies:

$$k_x = n_x \frac{\pi}{L}, \quad k_y = n_y \frac{\pi}{L}, \quad k_z = n_z \frac{\pi}{L}, \quad n_{x,y,z} = 1, 2, ... \tag{5.6}$$

(just like standing waves on a string, or modes of a drum). Note that $n = 0$ doesn't satisfy the boundary conditions, while $n < 0$ is already taken account of in the backward going part of the wavefunction. Hence the final solution is

$$
\begin{aligned}
\psi &= \frac{1}{L^{3/2}} \sin(k_x x) \sin(k_y y) \sin(k_z z) e^{-i\omega t} \\
E &= \frac{\hbar^2}{2m} \frac{\pi^2}{L^2} \left(n_x^2 + n_y^2 + n_z^2 \right) = \frac{\hbar^2 k^2}{2m}
\end{aligned}
\tag{5.7}
$$

There are many energy levels in the jellium box, with energies depending on their quantum numbers, n_x, n_y, n_z. The restrictions on k_x mean that we should redraw the curve in Fig. 5.3 as a series of points equally spaced in k.

What do we expect if we now fill our empty solid with electrons? Electrons are fermions and hence are indistinguishable if you exchange them. From the symmetry of the wavefunction it follows that only two electrons with opposite spin can fit in each level. Since we put in many electrons they fill up the levels from the bottom up. If we ignore thermal activation, i.e. take $T = 0$ K, this sharply divides the occupied and empty energy levels at a maximum energy called the Fermi energy, E_F.

Let us briefly estimate the typical separation of energy levels in real materials. Assuming a 1mm crystal, quantisation gives separations of 10^{-30} J = 10^{-12} eV. This can be written as a temperature using $E = k_B T$, giving T = 10^{-8} K. This shows that usually the separation of energy levels is very small; the average lattice vibration at room temperature has E = 300 K. So the levels are so close they blur together into a continuum.

Which electron energies are most important for the properties of jellium? When we try and move electrons so they can carry a current in a particular direction (say $+x$), they are going to have to need some kinetic energy. Without the applied field, there are equal numbers of electrons with $+k_x$ and $-k_x$ so all the currents cancel out. The electrons which can most easily start moving are those with the maximum energy already, so they only need a little bit more to find a free state with momentum $k_x > 0$.

Thus what is going to be important for thinking about how electrons move, are the electrons which are near the Fermi energy.

This is a typical example of how a basic understanding of a problem allows us to simplify what needs to be considered. This is pragmatic physics - otherwise theories get MUCH too complicated very quickly, and we end up not being able to solve anything. So if you really like solving problems exactly and perfectly, you aren't going to make it through the process of selection by intuition. A basic rule you should have is: first, build up an intuition, and then: go ahead and solve your problem based around that. If it doesn't work, then you know you need to laterally shift your intuition.

5.2 Density of States

We have found the energy levels for electrons in jellium, but to describe how the electrons move we will need to know how many electrons have any given energy. The number of electron states within a given range Δ is given by $D\Delta$ where the density of states is D. Particularly, as we discussed above, we need to know how many states there are at the Fermi energy. Very many physical properties which change with external conditions such as temperature or magnetic field, are acting by changing the density of occupied electron states at the Fermi energy. You have seen the first implication of quantum mechanics in jellium: that momenta can only take on specific values provided by the boundary conditions. This leads to the second implication of quantum mechanics for fermions in jellium: that at each energy value there are only a certain number of energy states available. Actually, we know the states occur at particular exact energies, but they are so close together we can never measure them individually (or not until we make the conditions very special). Instead we want to know how many states there are between two close energies, say E, and $E + dE$. To calculate this we will find it easiest to first find the number of states between k and $k + dk$, and then convert this to energies since we know $E(k)$.

We know each state is equally spaced in k every π/L, so the volume per state is $(\pi/L)^3$. Thus the number of states between k and $k + dk$ is the volume of the shell/volume per state (where we just take the positive values of $n_{x,y,z}$ because of the boundary conditions)

$$D(k)dk = \frac{4\pi k^2 dk}{(\pi/L)^3}\frac{1}{8} = \frac{k^2}{2\pi^2}dkL^3 \tag{5.8}$$

Unsurprisingly the number of states depends on the surface area since they are equally spaced throughout the volume, so $D(k) = k^2/2\pi^2$ per unit volume.

Each state can be occupied by either spin (up or down) so we multiply the density of k states by 2. Now we want to know how many states there are between two close energies at E and $E + dE$, so we differentiate the dispersion relation:

$$E = \frac{\hbar^2 k^2}{2m}$$
$$dE = \frac{\hbar^2 k}{m}dk = \hbar\sqrt{\frac{2}{m}}\sqrt{E}\,dk \tag{5.9}$$

So substituting into $2D(k)dk = D(E)dE$,
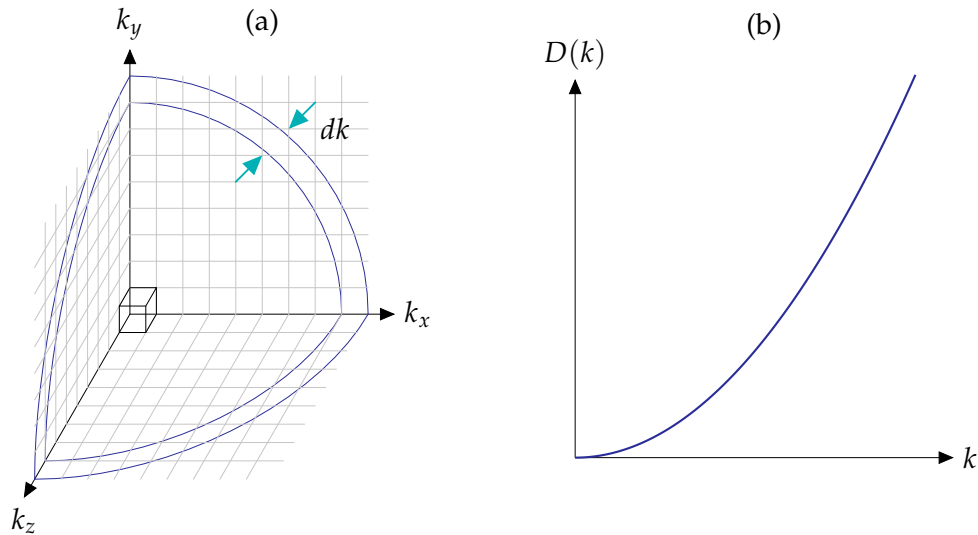
Figure 5.5: (a) Quantisation of k, (b) Density of states in k.

$$D(E)dE = \frac{V}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}\sqrt{E}\,dE \tag{5.10}$$

Often this is given per unit volume (by setting $V = 1$ in Eq. 5.10).

This is a typical example of calculating a property that will be useful to us. Remember we really want to know how many electrons have close to the maximum energy, the Fermi energy, given that we fill the levels up one by one from the lowest energy. So we counted the states, but because there are so many states we can approximate the actual number to the curve in Fig. 5.6.

To find the Fermi energy we first can calculate what the maximum k-vector of the electrons will be: Each of the total number of electrons we are dropping into the jellium, N_{tot}, has to be somewhere in the sphere of maximum radius k_F.

$$N_{tot} = 2\left(\frac{4\pi k_F^3}{3}\right)/8\left(\frac{\pi}{L}\right)^3 \tag{5.11}$$

$$k_F^3 = 3\pi^2 n \tag{5.12}$$

where n is the number of electrons/unit volume.

Now we can calculate what Fermi energy this gives:

$$E_F = \frac{\hbar^2 k_F^2}{2m} = \frac{\hbar^2}{2m}(3\pi^2 n)^{2/3} \tag{5.13}$$

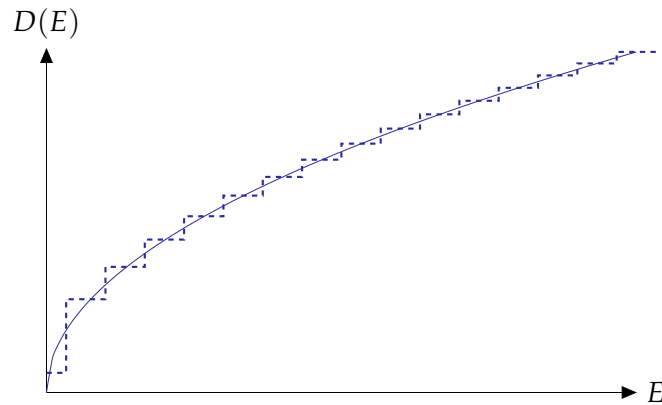Let us consider if this seems reasonable for a metal:

Figure 5.6: Density of states in energy. The solid line is our approximation while the dashed line is the actual density of states.

$n \sim 10^{29}$ m^{-3} (we see how we measure this later)

so $k_F \sim 10^{10}$ m^{-1} (in other words, electron wavelengths on the order of 1 Å)

so $E_F \sim 10^{-18}$ J ~ 7eV $\sim 10^4$ $K >> k_B T = 25$ meV , near room temperature.

These are very useful energy scales to have a feel of, since they are based on an individual electron's properties.

Remembering a little about our atomic elements, it is easier to remove a certain number of outer electrons, since they are bound less tightly to the nucleus. This is reflected in the electron density. We can rewrite $D(E_F)$ in terms of the Fermi energy by remembering that in one atomic volume there are z free electrons, so that $n \sim z/a^3$, and substituting gives:

$$D(E_F) = \frac{3z}{2E_F} \tag{5.14}$$

(density of states at Fermi energy per atom).

Thus for aluminium which has $z = 3$, and $E_F = 11.6$eV, $D(E_F) \sim 0.4$ eV^{-1} atom^{-1}.

5.3   Charge Oscillations

So now we know what happens when we put electrons into our jellium: the electrons fill up the potential well like a bottle, to a large energy. Now we want to know what happens as these electrons slosh around inside the bottle. We have to remember that as well as the uniform electron negative charge, there is a background fixed uniform density of positive charge from the ions. If the electrons are displaced by a small distance x, two regions of opposite charge are created:

The electric field produced between the two charged volumes is found from Gauss' theorem:

$$\oint \mathbf{E} \cdot d\mathbf{S} = \frac{Q}{\varepsilon_0} \implies \mathcal{E}A = \frac{nexA}{\varepsilon_0} \tag{5.15}$$
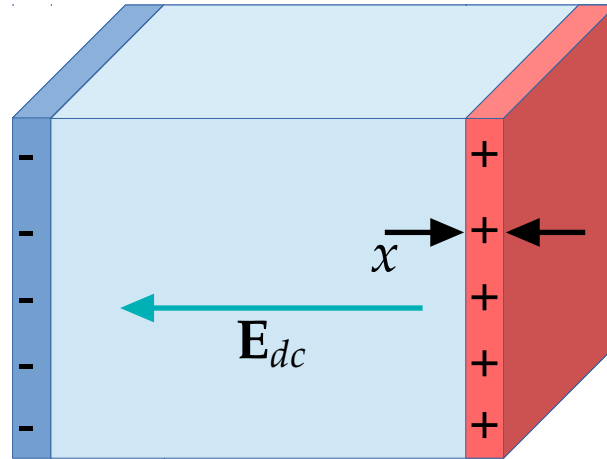
Figure 5.7: Polarisation from shifting electron distribution.

The electric field now acts back on the charges to produce a force on the electrons which are not fixed, so:

$$m\ddot{x} = -e\mathcal{E} = \frac{-ne^2}{\varepsilon_0}x \tag{5.16}$$

Since the restoring force is proportional to the displacement, this solution is easy - once again simple harmonic motion, with a "plasma frequency"

$$\omega_p^2 = \frac{ne^2}{\varepsilon_0 m} \tag{5.17}$$

An oscillation which emerges like this as a collective motion in a particular medium (here a gas of electrons), can often be thought of as a particle, or a '**quasiparticle**', in its own right. This may seem a strange way to think, but it is similar to the concept of electromagnetic fields in terms of photons. Because of the peculiar nature of quantum mechanics in which energy is transferred in packets, the photon quasiparticle can be convenient for the intuition. Actually, although we can discuss whether these quasiparticles 'really' exist, it seems that we can never decide experimentally. The charge oscillation quasiparticles which have an energy ω_p are called plasmons.

For "good" metals such as Au, Cu, Ag, Pt, the jellium model predicts that they are good reflectors for visible light ($\omega < \omega_p$). This breaks down - e.g. Cu and Au have a coloured appearance - which is a clue that the jellium model does not give a complete and adequate description for most materials.

We have some clear predictions for the way electrons fill up the jellium, which is our model metal. Because the electrons' underlying symmetry forces them to sit in different states, they are piled in higher energy states reaching very high energies. Also because they have a large density, they slosh back and forth very quickly; typically at optical frequencies. This optical property depends on all the electrons but now we are going to look at how the electrons move in the jellium to carry current.
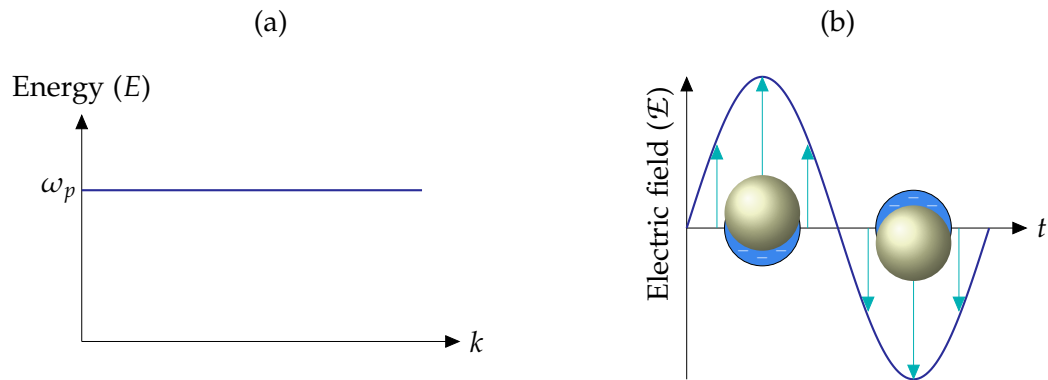
Figure 5.8: (a) Dispersion relation for plasmons. (b) Metal cluster undergoing SHM independent of λ.

5.4 Thermal Distribution

The electrons nearest the Fermi energy are important for thinking about the properties of the solid. Let's look at a first example. We have assumed that the electrons that fill up the jellium stay in their lowest possible energy states. However because the temperature is not zero, you know that they will have some extra thermal energy (see PHYS2024). The probability that an electron is excited to a higher state is proportional to $e^{-(\Delta E/k_B T)}$. Including the maximum state occupancy of two, this produces a blurring function as shown below in figure 5.9:
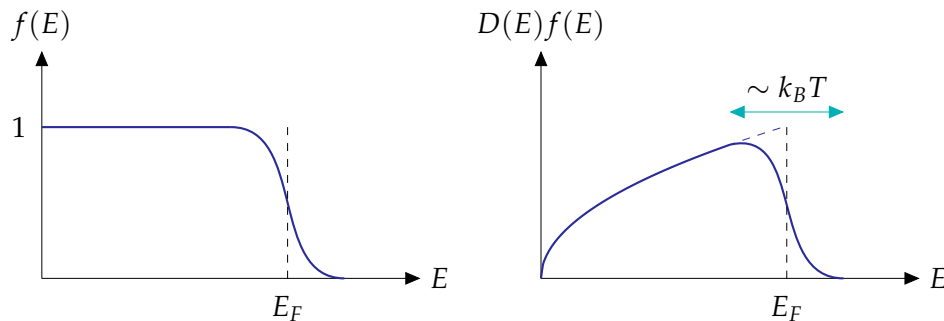
Figure 5.9: Thermal blurring of occupation of electron states at different T

The blurring function is called the Fermi-Dirac distribution; it follows from the fact that electrons are fermions, and that in thermal equilibrium they have a finite probability of being in an excited state.

$$f(E) = \frac{1}{e^{(E-E_F)/k_B T} + 1} \tag{5.18}$$

The blurring is much smaller than the Fermi energy and so the fraction of excited electrons is determined by the ratio of thermal energy and the Fermi energy: $k_B T/E_F$.

5.5 Electronic Transport

To first consider how we can move electrons around by applying an electric field to them, remember that $mv = \hbar k$. An individual electron on its own would accelerate due to the applied voltage:

$$m\frac{dv}{dt} = -e\mathcal{E} \tag{5.19}$$

What this says is that a constant electric field applied to an electron at a particular point on the dispersion relation will cause its k to continually increase.

Now we have our first picture of what a material would be like if it contained free electrons that could slosh around inside it. For applications we need to understand how electrons can be driven by voltages. We know the electrons will feel any electric fields that we apply to our solid. However we also can predict right from the start that unlike electrons in a vacuum, they will come to a stop again if we turn off the electric field. This is the essential difference between *ballistic* transport in which electrons given a kick just carry on going until they bounce into a wall, and *diffusive* transport in which electrons are always bouncing into things which can absorb some of their kinetic energy.
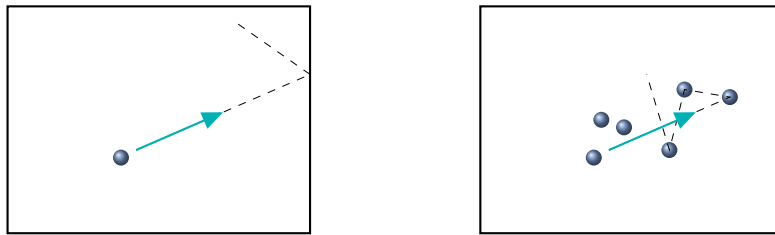
Figure 5.10: Ballistic and diffusive transport.

The electrons at the Fermi energy have an average velocity v_F given by $\frac{1}{2}mv_F^2 = E_F$, so:

$$v_F = \sqrt{\frac{2E_F}{m}} = \sqrt{\frac{2 \times 1.6 \times 10^{-19} \times 7}{10^{-30}}} \sim 2 \times 10^6 \text{ ms}^{-1}$$

about 1% of the speed of light (so very fast!). This is their quantum mechanical kinetic energy since it comes from dropping the electrons into a box that they can't escape from.

We are going to keep our model simple by ignoring for the moment, any details of *how* the electrons lose energy when they collide. The basic assumption is that the energy loss process is random. We will assume the most basic thing, that the speed of the electrons is subject to a frictional force, and this is determined by the time, τ, it takes the velocity to decay to $1/e$ of its value.

So for the whole group of electrons we will keep track of their average velocity, v_d

$$m\frac{dv_d}{dt} + m\frac{v_d}{\tau} = -e\mathcal{E} \tag{5.20}$$

The first term is the acceleration produced by the electric field while the second term is the decay of the energy due to inelastic collisions. What we have to note is that because the
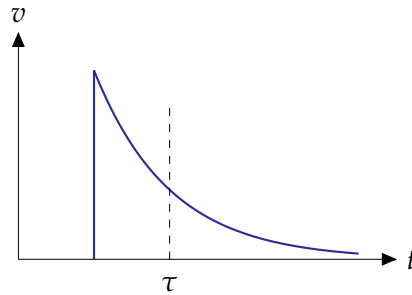
Figure 5.11: Decay in velocity of electrons.

scattering redistributes energy so fast, the drift velocity v_d is much smaller than the Fermi velocity v_F (waves vs. particles). Classically, 2τ is the effective acceleration time of electrons before a collision which randomises their velocity.

Now let us look at several cases:

5.5.1 Steady State

$$m\frac{dv_d}{dt} = 0 \tag{5.21}$$

so $mv_d/\tau = -e\mathcal{E}$, giving:

$$v_d = -\frac{e\tau}{m}\mathcal{E} \tag{5.22}$$

This drift velocity is proportional to the electric field applied, and the constant of proportionality is called the mobility (i.e. $v_d = \mu\mathcal{E}$) so that here:

$$\mu = -\frac{e\tau}{m} \tag{5.23}$$

which depends on the particular material. Because nv_d electrons pass a unit area every second, the current density

$$j = -nev_d = \frac{ne^2\tau}{m}\mathcal{E} \tag{5.24}$$

but remember $j = \sigma\mathcal{E}$ (i.e Ohms law!).

The conductivity, σ, is just the reciprocal of the resistivity ρ:

$$\sigma = \frac{1}{\rho} = \frac{ne^2\tau}{m} \tag{5.25}$$

We have calculated the resistance for our free electron gas, and shown that it produces Ohm's law. The relationship is fundamental even in more complicated treatments. To truly follow Ohm's law, the scattering time τ must be independent of the electric field applied, so we have hidden our ignorance in this phenomenological constant.

Let us estimate the scattering time for metals. If we take aluminium at room temperature, with a resistivity of 10^{-9} Ωm, then we find $\tau \sim 8$ fs ($\sim 10^{-14}$ s) which is extremely fast. For instance, it is about the same as the time an electron takes to orbit the nucleus of a hydrogen atom. As a result, when we apply 1V across a 1mm thick Al sample (difficult to do in a metal), the drift velocity produced is 5×10^{-2} ms^{-1} which is much less than the Fermi velocity.

The resistance is also extremely dependent on temperature - the resistance of most metals decreases strongly as the temperature is reduced, so that at liquid helium temperatures (T = 4K) copper is up to 10,000 times better conducting than at room temperature. This resistance ratio is a good measure of the purity of a metal.

Let us try and build an intuition microscopically for how conduction works in jellium. The extra drift velocity given to the electrons in the direction of the electric field, is an extra momentum which we can write in terms of an extra Δk.

From Eq. 5.22, we get:

$$mv_d = \hbar\Delta k = -e\mathcal{E}\tau \qquad (5.26)$$

So the electrons are shifted in k-space by Δk, which we can show graphically:
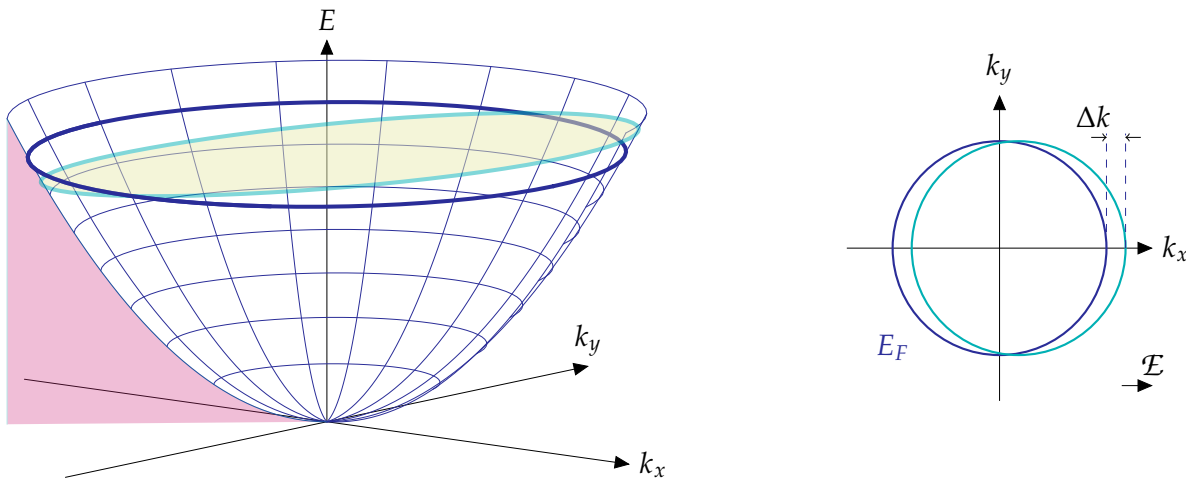
Figure 5.12: Shift in filled k-states when electric field applied in *x*-direction.

Each electron feels a force $F = \frac{dp}{dt} = \hbar\frac{dk}{dt} = -e\mathcal{E}$. Thus their k is constantly increasing, however they are rapidly scattered back to a different k after a time τ.

Because Δk is so much smaller than k_F, only two thin caps of k-states around the Fermi energy have changed their occupation. The electrons in these states are the ones which carry the current we measure. So we can see now why it was important to know how many states there are at this energy, $D(E_F)$, because this determines how many electrons are available to carry the current. When we remove the electric field, the Δk decays away and the original Fermi sphere is restored.

The electrons which carry the current have the Fermi velocity v_F, and so they travel a mean free path $l = v_F\tau$ before they collide and start being re-accelerated. For our typical metal, the mean free path is 300 nm, which is roughly a thousand time the interatomic spacing. This ties in reasonably with our assumption that the electrons are free, and not bound to a particular

atom. We will discover later that electrons can move freely in perfect periodic structures - scattering only occurs when this regularity is disrupted (e.g. defects). So our result so far is providing us with some justification for our initial assumptions.

5.5.2 Quantum Jellium (brief interlude)

Since we are looking at how far electrons move before they are scattered, we will ask what happens if try to make electrons pass through an incredibly thin wire. If we make our jellium box as small as the Fermi wavelength, $\lambda_F = 2\pi/k_F$, then the energy between the different electron states will become so large that scattering into them becomes impossible. Electrons now travel ballistically down the wire.
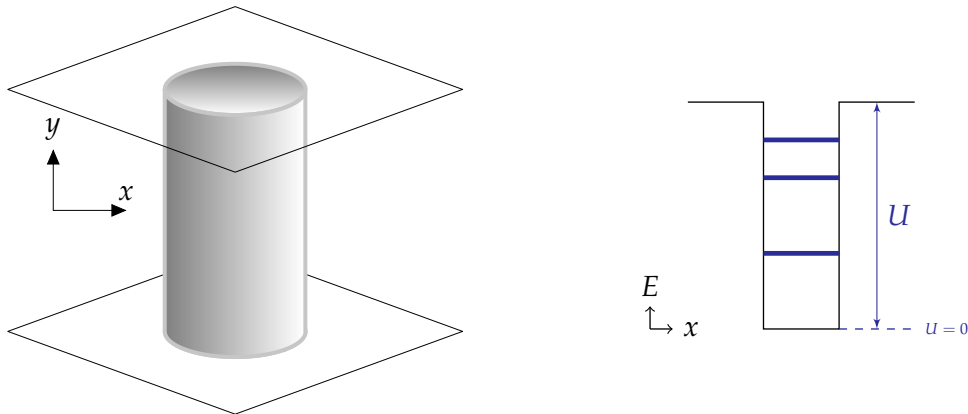
Figure 5.13: Confinement in a quantum wire connected to two conductive surfaces. Energy levels for $k_y = 0$

We can calculate what the resistance of this ultra-thin wire would be and we will see a surprising result, which links electronics to the quantum world. The wavefunction of our electron in this thin box is now given by a plane wave in only the y direction:

$$\psi = \frac{1}{\sqrt{L_y}}\xi(x,z)e^{ik_y y} \tag{5.27}$$

with:

$$E = \frac{\hbar^2 k_y^2}{2m} \tag{5.28}$$

Once again only states near the Fermi energy will be able to transport current, so we need the density of states in this one dimensional wire.

$$D(k_y)dk_y = \frac{dk_y}{(\pi/L)} \times \frac{1}{2} \tag{5.29}$$

which produces the density of states:

$$D(E) = \frac{m}{\pi\hbar^2 k_y} \tag{5.30}$$

per unit length.

As we discussed before, the electrons which can carry current are at the Fermi energy because the voltage applied lifts them into states which are not cancelled by electrons flowing in the opposite direction.
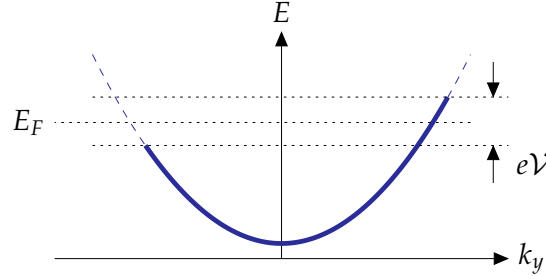
Figure 5.14: Filled states in 1D jellium when a voltage $\mathcal{V}$, is applied.

The number of uncompensated electron states is given by:

$$N = \Delta E \, D(E_F) = \frac{m}{\pi \hbar^2 k_F} L_y \Delta E \tag{5.31}$$

To calculate the current $I = Ne \, \mathcal{J}$, we need to calculate the flux $\mathcal{J}$ produced by the wavefunction for a plane wave. You saw how to do this in the quantum course, from the expression:

$$\mathcal{J} = -\frac{i\hbar}{2m} \left(\psi^* \frac{d\psi}{dy} - \psi \frac{d\psi^*}{dy} \right) = \frac{\hbar k_y}{m} \psi^* \psi = \frac{\hbar k_y}{m L_y} = \frac{v_F}{L_y} \tag{5.32}$$

which is nothing other than the electron velocity normalised to the probability density.

If we now substitute for the current:

$$\begin{aligned}
I &= D(E_F) e \mathcal{V} e \, \mathcal{J} \\
&= \frac{m}{\pi \hbar^2 k_F} L_y e^2 \mathcal{V} \frac{\hbar k_y}{m L_y} \\
&= \frac{e^2}{\pi \hbar} \mathcal{V} \tag{5.33}
\end{aligned}$$

So that within our approximations, the current is independent of both the length of the wire, and of the Fermi wavevector (and thus the density of electrons in the wire). Instead it depends simply on fundamental constants.

$$R = \frac{\mathcal{V}}{I} = \frac{h}{2e^2} = 12.9 \text{ k}\Omega \tag{5.34}$$

This is the quantum resistance of a single quantum wire, and is a fundamental constant of nature. Even more amazing for our simple calculation, when we measure experimentally very thin wires we indeed find exactly this resistance. It is so accurate that it is becoming a new way of defining current (instead of balancing mechanical and magnetic forces using currents through wires).

It applies to any thin wire in which the energies between the electron states which are produced by the diameter of the wire are larger than $k_B T$. For metals, the Fermi wavelength is very small, on the order of 0.1 nm, and so the wire needs to be this thin to act in the same way. This is a difficult wire to make a long length of, but in our approximation the length is not important. It is also possible to fabricate a wire from semiconductor materials, which as we shall see have a larger Fermi wavelength and can be made using standard microelectronics processing.

5.6   Checking the Jellium Model

From very basic principles we have derived some properties we expect for material with free electrons in it. All we needed was the electron density n, and the scattering time τ. We need this to understand even the very basics of what electrons are doing in a metal or semiconductor. But how do we know that we have captured the essence of electron behaviour in this incredibly simple model? We have seen that some of the optical and electrical properties seem quite sensible and agree with our experience. What other things can we test?

5.6.1   Hall Effect

One of the major guesses in our model so far has been the density of electrons. However there is an experiment which allows us to measure this value, in a simple and accurate way. It is not obvious that can we use a current to measure the number of electrons because $j = -nev_d$ means that a larger current can be caused by a larger n or v_d , or even by a carrier with a different charge. The trick is to devise a measurement in which the electron drift velocity cancels out. And for that we need to use a magnetic field as well.
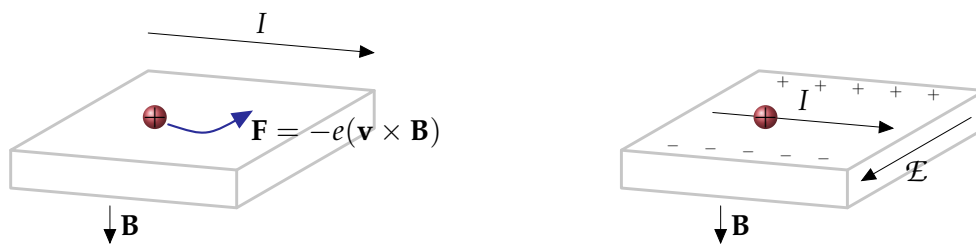
Figure 5.15: Hall effect in a conducting sheet with perpendicular magnetic field B (note we have taken a positively-charged carrier here).

The Hall effect was discovered by Edwin Hall in 1879 when he was a graduate student in the Johns Hopkins University. At that time, even the electron was not experimentally discovered so a clear understanding had to wait until quantum mechanics came into appearance. However from a basic analysis of the forces, it was possible to extract the charge density present in a conducting sheet.

He used a flat sheet of conductor so that conduction only occurs in a 2D plane. Although Hall originally used a metal plate, we now get even more extreme and strange behaviours if we use nanometre-thick sheets of electrons (similar to the quantum jellium wire) sandwiched inside semiconductors. We pass a current along this sheet, and then we apply a magnetic field perpendicular to the sheet.

i. The force on the electrons is $\mathbf{F} = -e(\mathbf{E} + \mathbf{v} \times \mathbf{B})$, so there is an extra force which because of the cross product, is perpendicular to both the B-field and the charges drift velocity (i.e. the direction of the current).

ii. So part of the current tries to flow across the sheet. Since it can't get out at the edges, it builds up there: extra positive charge on one side and negative charge on the other side.

iii. These charges now produce an electric field between the edges, which will build up until it exactly balances the force from the magnetic field. So in equilibrium, there will be no sideways force on the electrons.

Equating the lateral forces:

$$e\mathcal{E}_y = ev_x B_z \tag{5.35}$$

in which we can substitute for the drift velocity using $j_x = -nev_x$, giving:

$$\mathcal{E}_y = -\frac{j_x}{ne}B_z \tag{5.36}$$

This is often rearranged in terms of a Hall coefficient,

$$R_H = \frac{\mathcal{E}_y}{j_x B_z} = -\frac{1}{ne} \text{ in m}^3\text{A}^{-1}\text{s}^{-1} \tag{5.37}$$

This tells us the charge density as long as we know the charge of the current-carrying particle! Miraculously we have discovered a robust measurement since we know the current, magnetic field, and electric field. This is a great example of an experiment which is so cleverly designed that it gives you more than you expect. It is also a good example of serendipity - as Hall was not actually looking for this effect at all when he did the experiment. What he did do is understand that he had seen something interesting and interpret it within the framework of electromagnetism.

The Hall coefficient is an interesting test of the free electron model because it doesn't depend at all on the scattering time, τ. We can estimate the expected Hall coefficient using the number of electrons released for conduction by each atom, and the measured atomic densities.

Table 5.1: Hall coefficients for various metals.

Metal	Free electrons per atom	RH (10^{-11} m^3 A^{-1} s^{-1}) experiment	$-1/ne$ (10^{-11} m^3 A^{-1} s^{-1}) theory
Na	1	-23.6	-23.6
K	1	-44.5	-44.6
Cu	1	-5.5	-7.4
Ag	1	-9.0	-10.7
Mg	2	-8.3	-7.3
Be	2	+24.3	-2.5
Al	3	+10.2	-3.4
Bi	5	+54,000	-4.4

This table shows impressive agreement for the monovalent metals, given the simplicity of our assumption about the electrons! The group I elements (Na, K) are excellent, while being about

20% out for the transition metals (Cu, Ag). For the divalent metal Mg, it also gives reasonable agreement, however we head into trouble with Be, Al, and Bi since the sign of the coefficient is wrong. The experiment tells us that current in the material is being carried by positive charges! In the case of bismuth, something really strange is going on since we are four orders of magnitude out in our prediction, and many of these discrepancies exist in higher valences throughout the periodic table. The reason for this is something we will discover through this course and it holds the key to understanding the electronic properties of materials.

The Hall effect is also very useful for determining the appropriate carrier density grown into semiconductor materials. The strength of the Hall potential also is proportional to the strength of the magnetic field applied to the conducting strip, which in this context is known as a Hall Probe. A change in the magnetic field around the Hall probe produces a corresponding change in the Hall potential. Both of these parameters, the Hall potential and the strength of the magnetic field, can be measured by a Hall effect transducer or sensor. The instrument readily detects changes and sends signals to a monitoring device.

Many common applications rely on the Hall effect. For instance, some computer keyboards employ a small magnet and a Hall probe to record when a key is pressed. Antilock brakes use Hall effect transducers to detect changes in a car wheel's angular velocity, which can then be used calculate the appropriate braking pressure on each wheel. And Hall probes can be used (as in our labs) to measure very small and slow fluctuations in a magnetic field, down to a hundredth of a gauss (the unit of measurement for magnetic field strength, named after mathematician Carl Friedrich Gauss).

5.7   Beyond Jellium

We have been trying to understand the basic electronic properties of materials and started with the simplest system we could imagine, throwing electrons into a box (and remembering our quantum mechanics). We have got a long way in understanding metals, although we have now encountered a few problems.

We need to decide what it is that we have forgotten to take into account, and use it to correct our theory which didn't do too badly to start with. We have one more indication that some assumption made needs our attention, as there is a mismatch between our calculated optical reflection of metals and our experience. The plasma frequency for all realistic metallic carrier densities is much higher than the frequency of visible light, so all metals should reflect all colours. But experience tells us that metals like Cu and Au are reddish, so they don't reflect blue light as effectively as red light.

Both these deviations of our theory from our practical knowledge suggest that the way electrons move and the way they absorb optical energy is not based solely on their being free to wander through the crystal. We can try to guess what are the main fixups of our theory that we have to put in - electron behaviour will be affected by:

  i. the atomic structure of the material

  ii. the interaction between electrons

We know that some molten metals such as mercury conduct perfectly well, even though there is no long range crystal structure, so (i) can't be the only deficiency. However liquid metals are

much better examples of jellium than solids, so this gives an indication that crystal structure may be important.

We have also forgotten the fact that each electron moves not just in the confinement potential of the atomic lattice, but also feels the electric fields from all the other electrons as well. We have solved all our problems assuming that each electron can be treated independently, by assuming the other electrons move randomly and their effects average out.

The interaction between electrons is at its most extreme in the phenomenon of superconductivity. The pairing together of electrons, helped by lattice vibrations, allows them to cooperate with each other to form a new 'condensed' state. The lack of resistance for this state is produced by the lack of available levels for moving electrons to scatter into, near the energy of the condensed state. However correlated electron phenomena are generally not strong enough to survive to 300K, and so we will first concentrate on the effect of the atomic lattice on the movement of electrons.

Chapter 6

Nearly Free Electron Model

We have explored the jellium model of free electrons contained in a box, and found that it teaches us that free electrons in a solid form a metal from which we predict many very reasonable behaviours (electrical, optical, and also others not covered, such as thermal). We also found some discrepancies such as the electron density measured in the Hall effect, and we also failed to predict anything like insulating behaviour from these assumptions.

6.1 Quantum Mechanics in a Periodic Potential

We have seen above that electrons can be diffracted inside the regular arrangement of atoms inside a crystal (3.8)[1]. Our next job is to put together the two concepts of electron waves and electron diffraction to get a better idea of what electrons will do in a real crystal. And to give us an idea about how to do this we will start by going back to the description of electrons bound to atoms. Note that most of the derivations in this chapter deal with the 1-D (one dimensional) case only.

6.1.1 Electronic Potential

The first thing to do is look in finer detail at our picture of a solid. Consider first the potential energy of the outer electron on its own, in the combined potential of both the ion and the strongly-bound inner shell electrons. For the isolated ion with valence electrons, the outer electron is bound to the hydrogenic potential, and localised as expected.

When many ions bind together in a crystal, the atomic potentials are modified where they overlap, and the overall barrier height is reduced. The localised wavefunctions on each ion can now overlap, and the electrons become free to move between the ions if the barriers become low enough. Note: we have seen in chapter 1 that the outer electrons have a lower energy when the ions are in a crystal, providing the energy to bond the crystal together.

The potential that the electrons see is thus not flat as we assumed for Jellium, but is actually periodically modulated inside the crystal - it is the equivalent of electrons rolling around

---

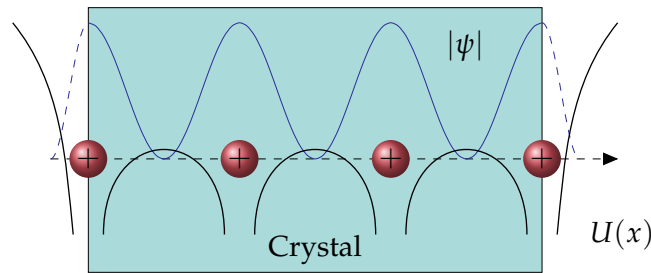[1]We will prove the condition for diffraction later on in this chapter.

Figure 6.1: Potential energy for outer electrons in a crystal.

inside an eggbox. To make a first attempt at understanding what happens to the electron energies, we are going to approximate this crystal potential with a sinusoidal potential (Eq. 6.1) as it turns out this produces much of the essential physics. We are also going to work in just one dimension for the moment because this will help make the mathematics simpler. We take the periodicity in the x-direction as a and define the vector $g = 2\pi/a$ (a reciprocal lattice vector), which will turn out to be very important.

$$
\begin{aligned}
U(x) &= 2U_0 \cos\left(\frac{2\pi x}{a}\right) \\
&= U_0\left(e^{igx} + e^{-igx}\right)
\end{aligned}
\tag{6.1}
$$

6.1.2 Solving the Schrödinger Equation

Having written down the Schrödinger equation for Jellium (Eq. 5.1), we now have to solve it with the new periodically modulated potential $U(x)$ instead of the Jellium box potential (in figure 5.1). Since we found the previous characteristics of the material produced by plane wave solutions e^{ikx} were not too bad, this relatively small periodic modulation must not change things too much. That means that it will still be useful to think in terms of plane waves of electrons with a particular k. The time-independent Schrödinger equation is given by (PHYS2003)

$$
\left[-\frac{\hbar^2}{2m}\nabla^2 + U\right]\psi = E\psi
\tag{6.2}
$$

where the left hand side has terms defining the kinetic and potential energy contributions of the wavefunction.

When the periodic potential is included, we know that each wavefunction is no longer a plane wave, but we can write each as the sum of plane waves possessing a different set of amplitudes, c_p ($p = 1, 2, 3, ...$). So we write for each particular wavefunction:

$$
\psi = \sum_{p=-\infty}^{+\infty} c_p\, e^{ipx}
\tag{6.3}
$$

The task is now to find the set of amplitudes, c_p, and the corresponding energy E for each electron wavefunction which is allowed in this periodic lattice. There are (as before) many

possible wavefunctions, and thus we should expect many possible energies to come from our solution.

If we just stick this wavefunction into Eq. 6.3 we get:

$$\left[-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}+U(x)\right]\sum_p c_p\, e^{ipx} = E\sum_p c_p\, e^{ipx} \tag{6.4}$$

We will use a (very common) trick to solve this equation, which depends on some of the beautiful properties of wavefunctions (however for the moment this may seem rather prescriptive to you). What we will do is to multiply the equation by e^{-ikx} and then to integrate both sides over all space (all this still keeps the equation true). The reason we do this is because when we multiply two plane waves together and then add up the result everywhere, we mostly tend to get nothing because the interferences tend to cancel. In other words, we use the useful result:

$$\int_{-\infty}^{+\infty} dx\, e^{-imx}e^{inx} = \delta(n-m) \tag{6.5}$$

This is called the orthogonality of plane waves and is illustrated in figure 6.2.
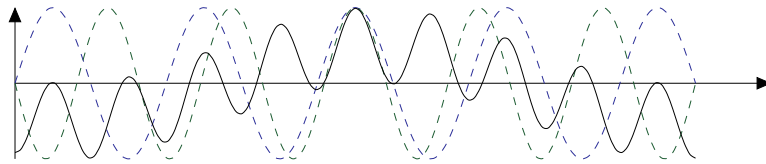
Figure 6.2: Orthogonality of plane waves.

So taking Eq. 6.4 $\times\, e^{-ikx}\int dx$, and swapping the order of sums and integrals gives us:

$$\sum_p c_p \int dx\, e^{-ikx}\left[-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}+U(x)\right]e^{ipx} = E\sum_p c_p \int dx\, e^{-ikx}e^{ipx} = Ec_k \tag{6.6}$$

The simplification on the right-hand side just comes from the orthogonality of plane waves, which we can also use to simplify the left hand side of the equation. When we substitute the potential in Eq. 6.1, this gives us:

$$
\begin{aligned}
Ec_k &= \sum_p c_p \int dx\, e^{-ikx}\left\{-\frac{\hbar^2(-p^2)}{2m}e^{ipx}+U_0e^{i(p+g)x}+U_0e^{i(p-g)x}\right\}\\
&= \sum_p c_p \int dx\, \left\{K(p)e^{i(p-k)x}+U_0e^{i(p+g-k)x}+U_0e^{i(p-g-k)x}\right\}\\
&= K(k)\,c_k + U_0\,c_{k-g} + U_0\,c_{k+g} \tag{6.7}
\end{aligned}
$$

Note we have defined the kinetic energy of an electron at k as:

$$K(k) = \frac{\hbar^2 k^2}{2m} \tag{6.8}$$

Crystalline Solids

which keeps the formula simpler later.

What this important equation shows is that a plane wave at k is now coupled together with plane waves at $k \pm g$, with the strength of this coupling being U_0 (which corresponds to the depth of the eggbox potential, due to the attraction of the ions in the lattice). We can rearrange this equation as:

$$(K(k) - E)c_k + U_0(c_{k-g} + c_{k+g}) = 0 \tag{6.9}$$

which should help you see that it is in fact a matrix of equations, for each $k = n_x \pi / L$. This equation (6.9) is still Schrödinger's equation but in a different form than you have met it before. To show what is happening, we can look at how the electron plane waves are coupled together on the dispersion curve for free electrons:
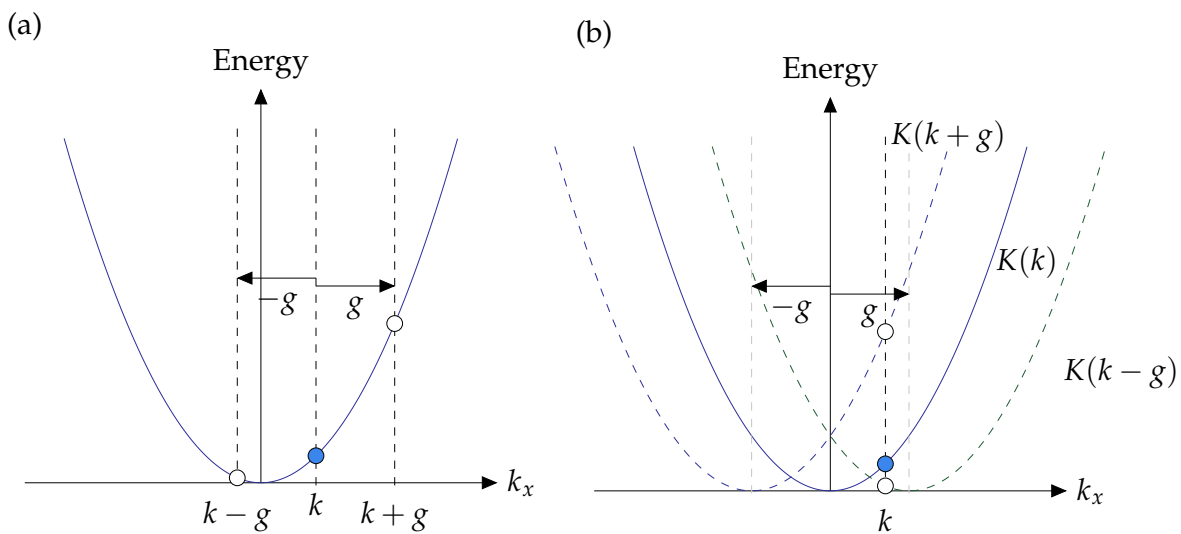

Figure 6.3: Coupling by the periodic ionic potential of plane waves at k and at $k \pm g$. (b) shows a different way of drawing it, so that all the waves coupled together are "folded back" to be at k.

You should see the similarity here with the diffraction grating which couples together optical plane waves in a similar way. Thus Eq. 6.9 has conservation of momentum built into it, with extra momentum being provided by the whole lattice in units of g. The solution to Eq. 6.9 is also dependent on the energies of the waves being coupled together, as we shall now see. We will tackle a few different cases to get an idea of how the matrix of equations work.

6.1.3 Non-degenerate Nearly-free Electrons

For a given value of k, we will consider the case when the corresponding free electron energy $K(k)$ is far apart from its value due to neighbouring reciprocal lattice vectors $K(k \pm g)$. For example, when $k = 0$, $K = 0$ and is far apart, in energy, from all other bands. This will reveal to us something about how the bands in this region are affected by the periodic potential.

This case begins by looking at very small deviations from the plane wave solutions. Note the terminology: since 'degenerate' means 'same energy' so 'non-degenerate' means 'different energies'.

Let's imagine that the corrugation on the bottom of the box is weak, so that it doesn't change the solutions for the wavefunctions of electrons very much. Because our free-electron model worked so well, we know this can't be too bad an approximation. In this case:

$$
\begin{aligned}
E_k &\approx \frac{\hbar^2 k^2}{2m} + \text{small corrections due to } U \\
c_k &\approx 1 \ , \ c_{k\pm g} \approx \text{small}
\end{aligned}
\tag{6.10}
$$

We then consider an unperturbed state with wavevector k', say, and energy $K(k')$. From this we then add the weak perturbation ($U_0 << E$). Hence $c_{k'} \approx 1$ and $c_{k \neq k'} << 1$. The significant coefficient $c_{k'}$ appears in three equations, which are a direct result of equation 6.9:

$$
\begin{aligned}
\left[K(k' - g) - E\right] c_{k'-g} &= -U_0(c_{k'-2g} + c_{k'}) \\
\left[K(k') - E\right] c_{k'} &= -U_0(c_{k'-g} + c_{k'+g}) \\
\left[K(k' + g) - E\right] c_{k'+g} &= -U_0(c_{k'} + c_{k'+2g})
\end{aligned}
\tag{6.11}
$$

From the first and third of these equations we obtain expressions for the coefficients:

$$
\begin{aligned}
c_{k'-g} &\approx \frac{-U_0}{K(k' - g) - E} \\
c_{k'+g} &\approx \frac{-U_0}{K(k' + g) - E}
\end{aligned}
\tag{6.12}
$$

Both are very small because U_0 is much smaller than the denominator. Combining these results with eq. 6.11 we obtain for the change of energy due to the perturbation:

$$
K(k') - E = U_0^2 \left[\frac{1}{K(k' - g) - E} + \frac{1}{K(k' + g) - E} \right]
\tag{6.13}
$$

The perturbed energy $E(k')$ of the electron is almost the original kinetic energy, $K(k')$, but shifted by an amount proportional to the square of the amplitude of the sinusoidal corrugation potential, and inversely proportional to the difference of the kinetic energy of electrons in the initial and diffracted waves. If you look at Fig 6.2, this difference in kinetic energies is large enough that the shift of the electron energy is quite small, so all our approximations work fine.

Much of this analysis is quite general: the closer together in energy that two unperturbed wavefunctions are before they are coupled by an extra potential, then the more they couple to each other and the more their original energies are altered.

The problem in our analysis comes when $K(k) = K(k + g)$ so that the denominators go to zero. This violates our assumption that only c_k is large, and E almost unchanged. Hence we have to separately look at the case that the diffracted electron wave energy is very close to the initial electron plane wave.
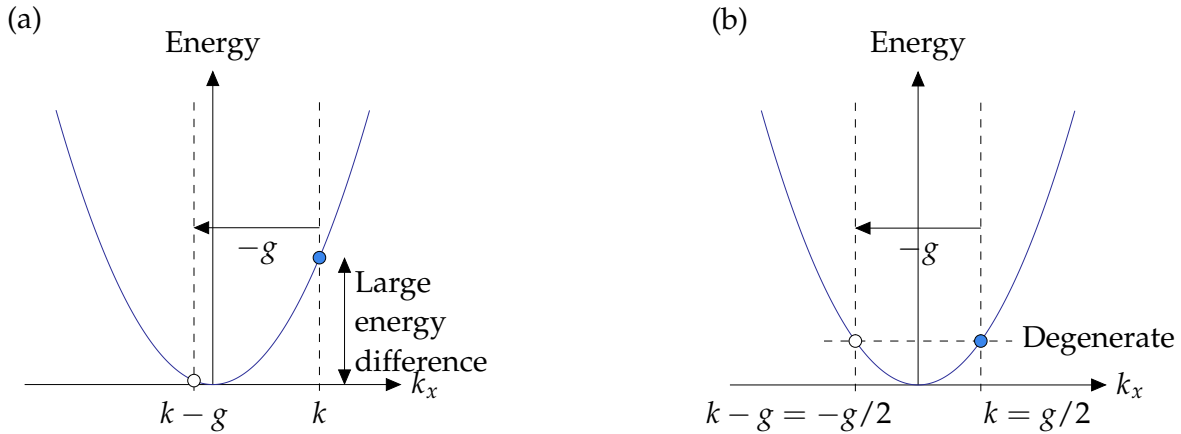
Figure 6.4: Initial and diffracted electron states for (a) non-degenerate and (b) degenerate cases.

6.1.4 Nearly-degenerate Nearly-free Electrons

In this case we are going to assume that the electron wave k has almost the same energy as the diffracted electron wave at $k - g$ (but not the same energy as any of the other diffracted electron waves, for instance at $k + g$. This is only to simplify the maths, and show the essential physics). In this case:

$$
\begin{aligned}
E_k &\approx \frac{\hbar^2 k^2}{2m} + \text{larger corrections due to } U_0 \\
&c_k, \ c_{k-g} \text{ are both large, } c_{k+g} \approx \text{small}
\end{aligned}
\tag{6.14}
$$

Now we have to write Eq. 6.9 for both c_k and c_{k-g}, neglecting small terms as before.

$$
\begin{aligned}
(K(k) - E)\, c_k + U_0\, c_{k-g} &= 0 \\
(K(k - g) - E)\, c_{k-g} + U_0\, c_k &= 0
\end{aligned}
\tag{6.15}
$$

In matrix form this gives:

$$
\begin{pmatrix} K(k) - E & U_0 \\ U_0 & K(k - g) - E \end{pmatrix} \begin{pmatrix} c_k \\ c_{k-g} \end{pmatrix} = 0
\tag{6.16}
$$

which has a non-zero solution only if the determinant of the matrix is zero:

$$
(K(k) - E)\,(K(k - g) - E) - U_0^2 = 0
\tag{6.17}
$$

or:

$$
E = \frac{1}{2}\left(K(k) + K(k - g)\right) \pm \sqrt{\left(\frac{K(k) - K(k - g)}{2}\right)^2 + U_0^2}
\tag{6.18}
$$

This is one of the key formulae for understanding the behaviour of nearly-free electrons, as it describes what happens when the initial and diffracted electron waves have nearly the same energy. But it looks quite complicated, so let us first look at the case when they have exactly the same energy.

6.1.5   Back-diffracted Degenerate Nearly-free Electrons

If initial and diffracted electron energies are equal, $K(k) = K(k - g)$, then $k = \frac{g}{2} = \frac{\pi}{a}$, as shown in figure 6.4.

The two new solutions from Eq. 6.18 for the electron energies are:

$$E_{\pm} = K\left(\frac{g}{2}\right) \pm U_0 = \frac{\hbar^2}{2m}\left(\frac{\pi}{a}\right)^2 \pm U_0 \tag{6.19}$$

Note that we start with two electron waves at the same energy, and we finish up with two solutions at different energies after we have included the effect of the corrugated ionic potential. The new wavefunctions are given by $c_{-g/2} = \pm c_{g/2}$, hence:

$$\psi_{\pm} = \frac{1}{\sqrt{2}}\left\{e^{ixg/2} \pm e^{-ixg/2}\right\} \tag{6.20}$$

The new energies are equally spaced around the original energy $K\left(\frac{g}{2}\right)$, and split by an amount $2U_0$. We can ask ourselves: why <u>physically</u> are there different energies for the two states at $k = \pi/a$? Firstly, the new wavefunctions are equal mixtures of $e^{i\pi x/a}$ and $e^{-i\pi x/a}$. These produce standing waves just as we saw for particles in a box though with a wavelength now related to the inter-atomic separation a, rather than the size of the crystal L.
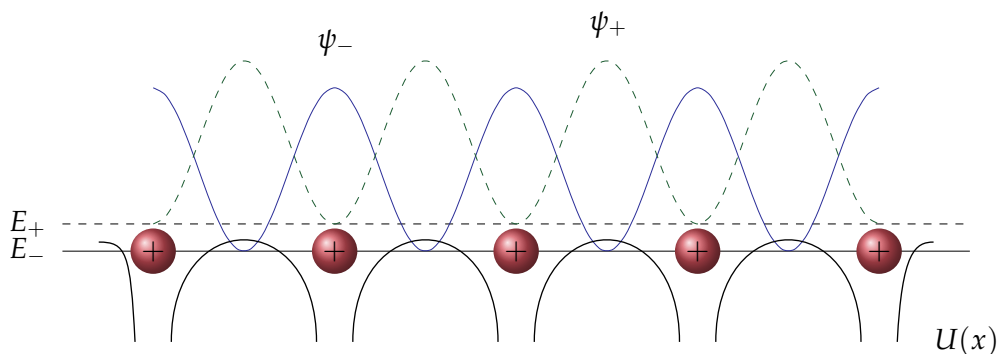
Figure 6.5: Ionic potential, and the spatial wavefunctions which are split in energy.

The two different combinations of these waves either give concentrations of electron density at the position of the positive ions, or in-between the ions. The standing wave with electrons most probable near the ions (ψ_-) has a lower energy because the electrons see a larger electrostatic attraction. The other state (ψ_+) has the electrons seeing very little of this attraction, and so has a higher energy. This gives the two split electron states we see.

Our starting point for understanding the behaviour of degenerate nearly-free electrons required that $K(k) = K(k - g)$. Now in general, $K(k)$ and $K(k - g)$ are only equal for some

reciprocal lattice vector when $|\mathbf{k}| = |\mathbf{k} - \mathbf{G}|$. This means that $\mathbf{k}$ must lie on a reciprocal lattice plane that bisects the line joining the origin of k to the reciprocal lattice point $\mathbf{G}$. This is equivalent to Bragg's law as discussed in section 3.2. As we have seen, a weak periodic potential most strongly affects (nearly) degenerate energy levels. We can therefore conclude that electrons with wave vectors that satisfy the Bragg condition will experience strong diffraction effects.

6.1.6 Understanding the Solutions for Nearly-free Electrons

What Eq. 6.18 tells us is that any two states which are scattered into each other by a potential U will produce new states which are now a mixture of the original ones E_1 and E_2 and with new energies:

$$E_\pm = \frac{1}{2}(E_1 + E_2) \pm \sqrt{\left[\frac{1}{2}(E_1 - E_2)\right]^2 + U_0^2} \tag{6.21}$$

The new states are centred around the average energy of the original two states, but are split apart by an amount which is the mean square average of their initial separation and the strength of the scattering.

Note that when the potential is turned off ($U = 0$) in Eq. 6.18, the energies go back to the two individual free-electron solutions, as they should do (sanity check).
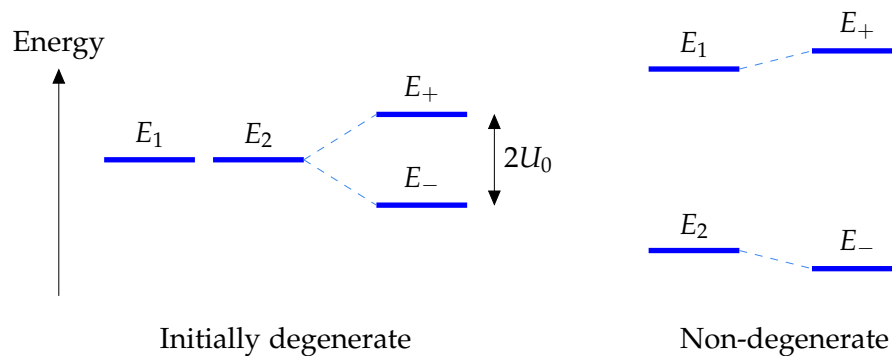
Figure 6.6: New states from coupling energy levels $E_{1,2}$ by a potential U.

6.1.7 Coupled Solutions

We are now in a position to look at the new dispersion relation for our real solid, by graphing Eq. 6.18.

The most important feature for you to notice is that for the first time we have built a model which says there is a range of energies which cannot be reached: *energy gaps*. Electrons can only adopt energies in certain ranges: *the energy bands*. From this derives all the vast functionality of modern electronics! We also see that if the free electron energies of k and $k - g$ are too far apart then the diffraction of the electron waves does not affect them much.

By adding extra curves of the free electron energies centred at each reciprocal lattice point ($k_x = +g$ and $-g$ below) this is easier to see. This is called the repeated zone scheme.

Figure 6.7: Dispersion relation in the "extended" zone scheme for nearly-free electrons (thick) and free electrons (dashed).
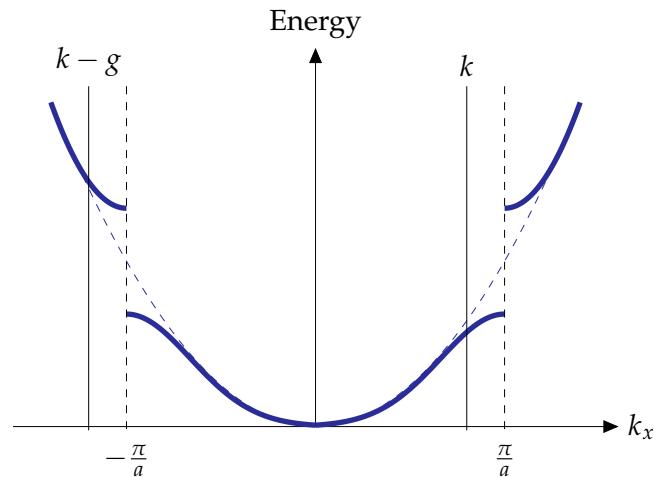
We can justify this shift in momentum by any multiple of $\hbar g$ because it corresponds to the fact that the lattice can always supply momentum of this amount, so we can never determine any k to better than $k \pm g$. This amount of momentum that the lattice freely adds or removes is so important that it is called the "reciprocal lattice vector", $g = 2\pi/a$, and you should see that it is defined in a similar way as the wavevector $k = 2\pi/\lambda$.

In this scheme, energy levels which are at the same k are able to couple with each other, since by repeating the plot centred at each reciprocal lattice point we have included all the possible lattice scattering. Only when $k_x \sim g/2 = \pi/a$ and the energies of the free electron states become comparable does the diffraction become significant, and produce an energy gap. The point $k = \pi/a$ is extremely important as the forward- and backward-going waves are equally mixed there, and the energy gap is at its smallest. This edge is called the Brillouin zone edge and we have found it in the x-direction. In fact the Brillouin zone corresponds to the Wigner-Seitz cell of the reciprocal lattice.
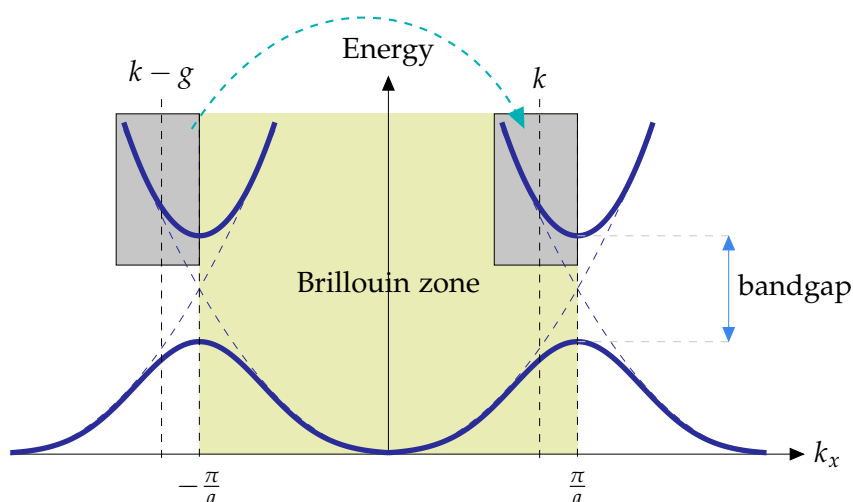

Figure 6.8: Dispersion relation for nearly-free electrons in the "repeated" zone scheme.

6.2 Approximations Revisited

Because we did not use the full mathematical framework for solving what happens when we throw our electrons into a box which has a corrugated bottom, we should discuss briefly what we might have left out.

1. We assumed that each wavefunction which is a solution for an electron inside this box can be written as a sum of all plane waves (Eq. 6.3). This turns out to be generally true, but sometimes not the most useful way to solve accurately for the electrons inside a crystal. One important fact we have not really used is that any solution for the electron density of an electron state should look exactly the same if we shift it by a lattice vector (**R** is any lattice vector):

$$|\psi(\mathbf{r} + \mathbf{R})|^2 = |\psi(\mathbf{r})|^2 \tag{6.22}$$

   This means that the wavefunctions must be the same, up to an arbitrary phase factor θ_T:

$$\psi(\mathbf{r} + \mathbf{R}) = e^{i\theta_T}\psi(\mathbf{r}) \tag{6.23}$$

   It turns out (Bloch's theorem, we will prove this in the following chapter) that the wave-functions can be written as the product of plane wave and a periodic function:

$$\psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k}}(\mathbf{r}) \tag{6.24}$$

   where $u_{\mathbf{k}}(\mathbf{r} + \mathbf{R}) = u_{\mathbf{k}}(\mathbf{r})$ repeats exactly the same for each unit cell.

   Effectively what this wavefunction does is to acknowledge the fact that not all k are coupled together in the new electron states, but only plane waves which differ by a wavevector g.

2. We assumed that the potential U was sinusoidal, which is not true. What we really need to do is calculate the left-hand side of Eq. 6.6 using the correct $U(x)$:

$$\int_{-\infty}^{\infty} dx\, e^{-ikx}\, U(x)\, e^{ipx} = \widetilde{U}(p - k) \tag{6.25}$$

   The function is called the "Fourier Transform" of $U(x)$, and it basically says how much lattice momentum that a crystal can provide. The more localised is the potential around each ion, the larger the range of momentum that the lattice can provide to the electron. You should also remember that the lattice can only provide momentum in units of $\hbar 2\pi/a$.

3. We solved the problem in only a linear chain of atoms (1D) rather than a solid crystal (3D). Of course now the lattice can diffract into a range of other directions, rather than sending the electron waves straight back the way they came in. This complicates the problem, but changes rather little of the physics. Perhaps the most important thing to see is that if the lattice constant is different in different directions (i.e. $a \neq b \neq c$), then the amount of momentum the lattice supplies can also be different in different directions: $a^* = 2\pi/a \neq b^* = 2\pi/b \neq c^* = 2\pi/c$. It turns out that the lattice can contribute any

general reciprocal lattice vector $\mathbf{G}$ to the electron wave $\mathbf{k}$, where $\mathbf{G} = h\mathbf{a}_1^* + k\mathbf{a}_2^* + l\mathbf{a}_3^*$. We shall see how this can make a difference to the electronic properties of a material below.

4. Finally, we assumed that when we corrugate the potential at the bottom of the box containing the electrons, they still remain basically free. In chapter 7 we will solve the whole problem again (as it is so important) but from the completely opposite point of view that the electrons start out completely localised at each ion, and then can hop from one to the next. What should reassure you is that the properties that emerge are consistent with the *nearly-free electron* model.

6.3 Electronic Properties

So far we have given the quantum mechanical solution for plane waves in the crystal. From the corrugated potential energy profile which is periodic in a crystal lattice, we have found the different energies of plane waves which can exist.

Next to follow the implications we have to fill up these electron states with electrons (as before). We still have the quantisation from the size of the crystal which gives the density of allowed plane wave k-states. If the density of free electrons is enough to produce a Fermi energy which is much larger or smaller than the middle of this energy gap, very little changes and we have a metal like jellium (Fig. 6.9 a,b).
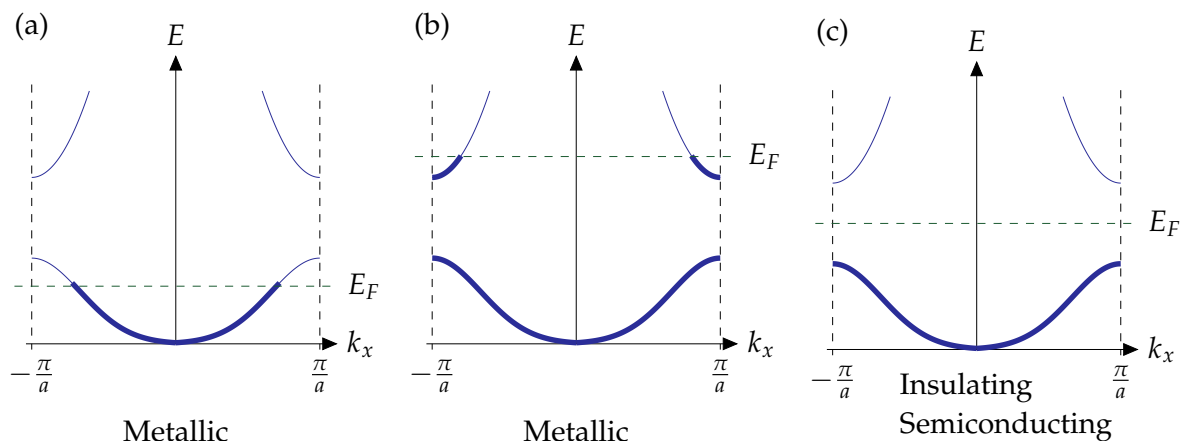
Figure 6.9: The nearly-free electron model for different electron densities in the "reduced" zone scheme.

6.3.1 Filling of Bands

The number of electron states in each band is simply found from the fact that they are spaced by $\Delta k = \pi/L$, hence the total number in each band,

$$2\frac{\pi/a}{\pi/L} = 2L/a = 2N \tag{6.26}$$

where N is the number of unit cells in the sample, and 2 comes from the property that an electron of each spin can fit into each energy level. Note: We only consider positive k

again in this calculation because just as in our original counting of the states in section 5.2, the boundary conditions we have used mean each $k = n\pi/L$ contains both forward- and backward-going plane waves. This means that if only 1 electron is released by the atoms in each unit cell, then these electrons will exactly half-fill the lowest band. So this crystal will be a free-electron metal (as we found for Na and K). Similarly for 3 electrons released per unit cell, the material will be a metal (Al), although not all the electrons can participate in conduction.

A huge contrast to metallic behaviour occurs if the electron density is such that the Fermi energy is near the midgap energy. In this case if 2 or 4 electrons are released by the atoms in each unit cell, then the electrons exactly fill all the lower energy states (called the valence band states) as in figure 6.9c. When we try and accelerate the electrons using an electric field across the material, they have no energy states to go into. They can't move and the material is an insulator.
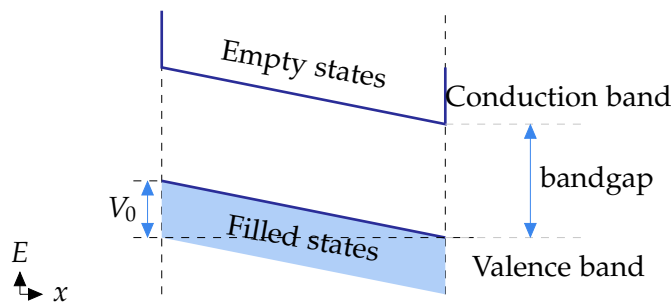

Figure 6.10: Potential energy of electrons near E_F as a function of x with an electric field applied

The electrons cannot move unless we give them much more energy, at least equal to the band gap (normally denoted E_g and equal to $2U_0$ in the calculations above). Even the best calculations have a hard time getting the exact size of the band gap E_g quite right, and often we just take it from experiment. This energy ranges from a few meV up to several eV. If the band gap is large then the material is always insulating and often transparent, e.g. silica SiO_2, sapphire Al_2O_3. If the band gap is small enough we call the material a semiconductor, and it becomes very useful for electronics. The reason is that by applying electric fields (whether with voltages on metallic contacts, or using light) we can influence the movement of electrons which have energies close to the band gap edges. And thus we can controllably turn the material from a metal into an insulator. This electronic functionality is at the heart of semiconductor electronics (see chapter 9).

Having started from first principles using the quantum mechanics you know, we have got to the stage where the properties of metals, insulators and semiconductors emerge. This is one of the most dramatic basic results of solid state physics. Essentially the formation of forbidden bands of energy for electrons arises because the periodic lattice strongly diffracts electron waves which have almost the same wavelength as the lattice period. The resulting standing waves have different energies because they 'feel' the attractive ion cores differently.

6.3.2 Electron Behaviour

Electrons in the states around the bandgap are especially important to understand their properties and control the electronic functionality of the solid. So we will now focus on just this

part of the dispersion relation, rather than the full equation we derived Eq. 6.21.

Electrons in the conduction band have a dispersion relation which looks very similar to the free electrons in jellium. The curvature of the dispersion is related to the mass of the electron by:

$$\frac{d^2E}{dk^2} = \frac{\hbar^2}{m} \implies m = \hbar^2 \left[\frac{d^2E}{dk^2}\right]^{-1} \tag{6.27}$$

Heavier particles have less steeply-curved dispersion relations since their momentum is larger for the same energy.
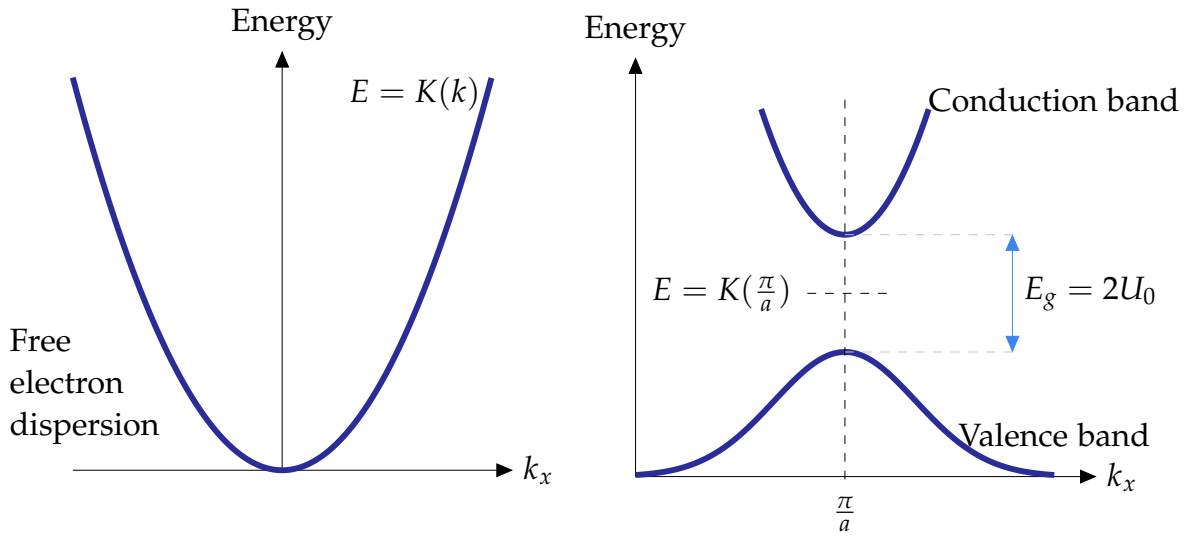
Figure 6.11: Dispersion relation of a) jellium b) nearly free electrons.

The curvature of the conduction band dispersion is also quadratic near to $k = \pi/a$. So we can write $k = \pi/a + \kappa$, and substitute in Eq. 6.21,

$$E_{\pm}(k) = \frac{1}{2}\left[K(\kappa + \frac{\pi}{a}) + K(\kappa - \frac{\pi}{a})\right] \pm \frac{E_g}{2}\left[1 + \left(\frac{K(\kappa + \frac{\pi}{a}) + K(\kappa - \frac{\pi}{a})}{E_g}\right)^2\right]^{1/2} \tag{6.28}$$

If we do a binomial expansion near the bandgap:

$$E_{\pm}(\kappa + \frac{\pi}{a}) = \frac{\hbar^2\pi^2}{2m_ea^2} \pm \frac{E_g}{2} + \frac{\hbar^2}{2m_e}\kappa^2\left[1 \pm \frac{2\hbar^2\pi^2}{m_ea^2E_g}\right] \tag{6.29}$$

where we have explicitly notated the normal electron mass in vacuum, m_e. Thus although the dispersion is quadratic, it now has a different curvature around the conduction and valence band edges which implies that the mass of the electrons is now:

$$m^* = m_e\left[1 \pm \frac{2\hbar^2\pi^2}{m_ea^2E_g}\right]^{-1} = m_e\left[1 \pm \frac{\left(\frac{\hbar^2(2\pi/a)^2}{2m_e}\right)}{E_g}\right]^{-1} \tag{6.30}$$

So the effective mass of the electron is modified by the ratio between the bandgap and the kinetic energy of the free electron wave at twice the Brillouin zone edge.

The effect of putting the electron in the periodic potential is to 'dress' it with scattered states which mean that it has a different inertia. Note that these electrons are often much lighter than the free electron mass. This is an example of a frequently used concept in physics called a quasiparticle. Although you have come across them before in the context of plasmons, we will still call these quasiparticles 'electrons'. It is important to remember that they have different properties to electrons in free space.

Because the conductivity of a material is simply written $\sigma = ne^2\tau/m^*$ (see Eq. 5.25) in terms of the particle mass, this quasiparticle behaviour has repercussions through most of the properties we calculated previously. The mass of the electrons in this estimate of the conductivity is the effective mass, and not the vacuum electron mass.

An even more extreme example of a quasiparticle is found in the valence band. When an electron is excited from a full valence band to an empty conduction band, it leaves behind an empty electron level. We can view this as a positively-charged hole left behind in the sea of electrons. As an analogy consider a bottle almost full of liquid so only a small air bubble remains. The natural way to describe what happens when the bottle is tilted is in terms of the air bubble moving rather than the much more complex motion of the liquid. If we do this we start to understand why the Hall coefficient which was supposed to measure the density of electrons might come out with a positive and not a negative sign. It is because the charge carriers can be positive holes near the top of the valence band.

This empty electron state (= 'filled' hole state) has a:

Wavevector, $k_h = -k_e$: The total wavevector of all the electrons in the full band is zero since each electron at $+k$ is compensated for by an electron at $-k$. Removing an electron with wavevector k_e leaves the hole with a wavevector $k_h = -k_e$.

Energy, $E_h = -E_e$: The electron previously in the valence band had an energy E_e so removing it leaves a hole with energy $-E_e$.

So the hole has the opposite momentum and energy to the electron that was previously there.
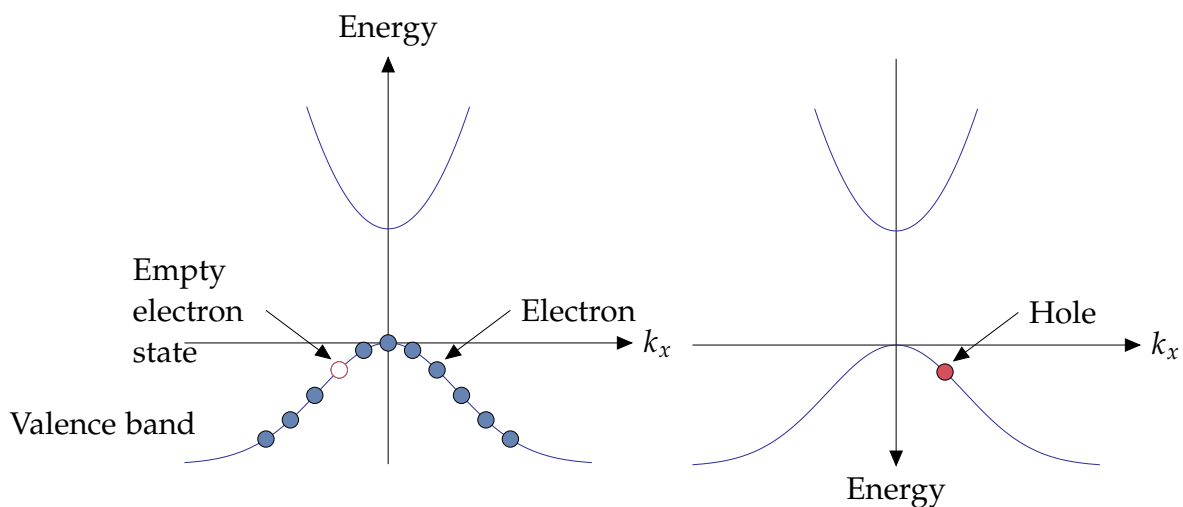

Figure 6.12: Dispersion relation of empty electron states and equivalent filled hole states in valence band.

Mass: of the quasiparticle is proportional to the curvature of the dispersion relation. The mass of the electrons at the top of the valence band is negative (!) but the curvature of the hole dispersion is opposite so the mass $m_h^* = -m_e^*$. Generally holes are heavier than electrons in the conduction band, which is not surprising since many electrons need to move for the hole to itself move.

Charge: Since an electron has been removed, the hole looks like a positive charge. When we now do the Hall effect measurement, the voltage applied will accelerate the holes in the opposite direction to the electrons (i.e. in the same direction as the current). As before, the lateral Hall field developed must cancel out the sideways force from the magnetic field, so $E_y = -v_x B_z$. However this velocity is in the opposite direction to the electrons, so the opposite lateral field will be produced, and unexpectedly be measured as a positive Hall coefficient.

The current carried by a material is now the sum of the electron and hole contributions: they are accelerated in opposite directions by a voltage applied, but contribute current in the same direction. Thus:

$$j = ne\mu_n \mathcal{E} + pe\mu_p \mathcal{E} \qquad (6.31)$$

$$\sigma = ne\mu_n + pe\mu_p \qquad (6.32)$$

This produces a Hall coefficient (remember NOT resistance):

$$R_H = \frac{1}{e} \frac{p\mu_p^2 - n\mu_n^2}{(p\mu_p + n\mu_n)^2} \qquad (6.33)$$

Thus if the hole concentration $p > n$, the Hall coefficient can be positive (remember Be).

Holes are very important quasiparticles in semiconductors because they behave so differently to electrons. The interaction between electrons and holes is the key to many properties of electronic devices.

6.4 Periodic Potential in Two and Three-dimensions

Our current understanding of how electrons behave in a weak periodic potential is readily extended to higher dimensions. This however leads to energy band structures with significant geometric complexity that are not easily visualised as a projection onto a single plane. It is often more important to determine the structure of the dispersion relation $E(\mathbf{k})$ in the immediate vicinity of the Fermi wavevector $\mathbf{k}_F$. In two and three-dimensions the divide between contiguous occupied and unoccupied energy states is referred to as the Fermi circle and Fermi surface, respectively. Often the term Fermi surface is used collectively.

For a two dimensional isotropic lattice, we can construct the Fermi surface by first considering that the periodic potential defines both the lattice and the corresponding reciprocal lattice points. We first begin by assuming that the periodic potential is sufficiently weak that it does not significantly alter the free electron properties. The Fermi surface therefore has a circular form (See figure 5.3, for example). In two dimensions, each unit cell contains a single atom

and therefore has room at most for two electrons. The density of states is then: $a^2/2\pi^2$. If k_{Fn} is the wavevector corresponding to n electrons per atom, then we can write:

$$\frac{a^2}{2\pi^2}\,\pi k_{F1}^2 = 1$$
$$\frac{a^2}{2\pi^2}\,\pi k_{F2}^2 = 2$$
$$\frac{a^2}{2\pi^2}\,\pi k_{F3}^2 = 3 \tag{6.34}$$

If we construct the Fermi circles for each of the above cases, we see that for k_{F1}, the Fermi circle lies within the first Brillouin zone where as for k_{F2} and k_{F2}, it intersects the zone boundary. If we then increase the magnitude of the periodic potential so that energy gaps arise, Fermi circles for k_{F2} and k_{F2} can no longer remain continuous as they cross the zone boundary. Any point that is infinitesimally close to, but just inside of, the first zone boundary, is situated at a different energy than that of an adjacent point just outside of the boundary, inside the second zone. The two states differ in energy by the energy band gap. A weak periodic potential therefore distorts the energy contours and draws 'necks' towards the zone boundaries.
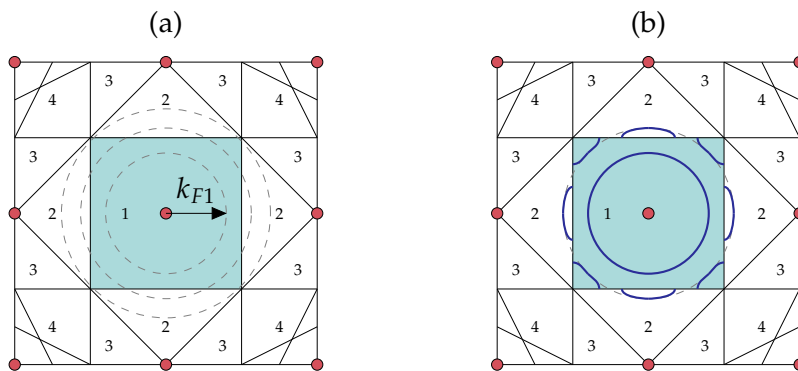

Figure 6.13: Two dimensional reciprocal lattice showing (a) free electron Fermi circles that intersect the first Brillouin zone boundary, (b), Deformation of the free electron circle near the Brillouin zone boundary due to the periodic potential (only k_{F1} and k_{F2} are shown, for clarity).

Chapter 7

Tight Binding

When we were discussing covalent bonding we talked about two atoms coming together to form a molecule. However most molecules contain more than two atoms and crystals contain enormous numbers of atoms. Let us consider what would happen if we had three atoms in a molecule. As this problem is even more complicated than the two atom case we will again use the linear combination of atomic orbitals approximation. Thus

$$\Psi_{\text{molecule}} = a\psi_{\text{atom}}^{(1)} + b\psi_{\text{atom}}^{(2)} + c\psi_{\text{atom}}^{(3)} \tag{7.1}$$

where a, b, c are constants to be determined from the Hamiltonian. As you will prove in the problem sheets the eigenstates are given by $(a, b, c) = (1, \sqrt{2}, 1)/2$, $(1, -\sqrt{2}, 1)/2$ and $(1, 0, -1)/\sqrt{2}$. As in the case of two atoms we would expect that the lowest energy eigenstates will concentrate electron density between the atoms. Thus after some thought it should be obvious that the lowest energy orbital is that given by $(a, b, c) = (1, \sqrt{2}, 1)/2$ i.e. when constructive interference of the atomic orbitals occurs between the atoms. It is also obvious that the highest energy orbital will be the one that has maximum destructive interference between the atoms i.e. $(1, -\sqrt{2}, 1)/2$. The energy of the eigenstates and schematic representations of their probability distributions are shown in Fig. 7.1.

In the tutorial problems for this course you will calculate the molecular orbitals for 4 and 6 atoms molecules. In general it is clear that the lowest energy molecular orbital is always the one with all the atomic orbitals in phase (this orbital is lower in energy than the atomic orbital by the overlap energy), the highest energy orbital is always where the phase of the atomic orbitals changes by π from an atom to it's neighbour (this orbital is higher in energy than the atomic orbital by the overlap energy), and the number of molecular orbitals equals the number of atoms. Therefore the energy spacing (note: energy per unit cell) between orbitals is inversely proportional to the number of atoms, N. When N is big enough the fact that the molecular orbitals have discrete energies becomes unimportant and instead we can treat the molecular orbitals as filling up a band of energies. The thought processes that are set out above are called the "tight binding model" of electronic energy band formation in solids. We will now look at this model in more detail.
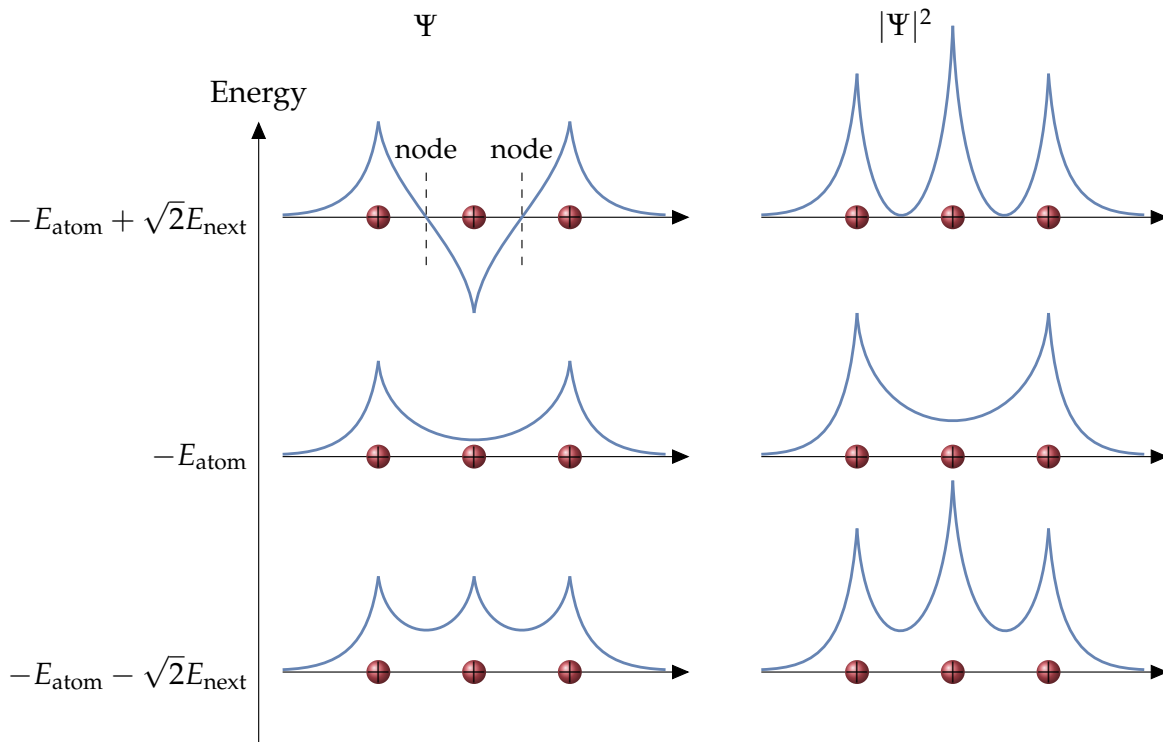
Figure 7.1: The three molecular orbitals of a 3 atom hydrogen molecule.

7.1 Description of Tight Binding Model

The basic approximation behind this model is that the effect of the other atoms in a solid on the electrons of a particular atom is only a weak perturbation. This is clearly true in the case of atoms that are far apart. Therefore another way of stating this approximation is that the probability of electrons on an atom being at a distance comparable with the separation between atoms in the solid should be small i.e. that the atom's electrons are tightly bound to that atom. Yet another way to look at this approximation is that it is just a restatement of the linear combination of atomic orbitals approximation. i.e. that the electron states in a solid are well approximated by sums of atomic orbitals.

Another extremely important approximation in the model is that it ignores the details of the coulomb interactions between electrons i.e. it treats the electrons as if they did not interact and therefore is a single electron model. In many cases this is a surprisingly good approximation however it means that this model would not predict iron to be ferromagnetic.

Now as physicists we want to make quantitative predictions with our models. However consider the problem of writing down wavefunctions consisting of vectors with 10^{26} numbers in them. Not really practical. However thankfully this is not necessary as we are saved by an extremely powerful theorem called Bloch's theorem.

7.2 Bloch's Theorem

This is based on the translational symmetry of a crystal lattice. That is if a crystal is moved in space by a lattice vector then it looks identical to before it was moved. The theorem states that

the wavefunctions, $\psi(\mathbf{r})$, of electrons in crystals always have the property that if we observe the wavefunction about a lattice point, $\mathbf{R}$, it is the same as about the origin except for a phase shift which is given by multiplying the wavefunction by $e^{i\mathbf{k}\cdot\mathbf{R}}$, i.e.

$$\psi(\mathbf{r} - \mathbf{R}) = \psi(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{R}} \tag{7.2}$$

Another way of stating this is that the wavefunction can always be written in the form:

$$\psi(\mathbf{r}) = u(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}} \tag{7.3}$$

where $u(\mathbf{r} + \mathbf{R}) = u(\mathbf{r})$ for all lattice vectors $\mathbf{R}$, i.e. that is the same at every lattice point. A number of different proofs can be given for this theorem. What I will present here is not a totally rigorous proof however it contains all the main points of a complete proof[1]. For a complete proof see the appendix or Ashcroft and Mermin pages 134 and 137. Consider an eigenstate $\psi(\mathbf{r})$ which is not of the form above but instead has a magnitude which is different at one or more lattice points. Now from this eigenstate we can make another different one for every lattice point by displacing the wavefunction to be centered about this lattice point. The energy of the eigenstates is given by the integral of $|\psi(\mathbf{r})|^2$ multiplied by the potential due to all the atoms in the crystal, which is periodic and thus all the eigenstates would have to have the same energy. Thus in an infinite crystal we would have to have an infinite number of eigenstates with the same energy. This situation is absurd and thus we know for certain that any eigenstate must have the property that $|\psi(\mathbf{r})|^2 = |\psi(\mathbf{r} + \mathbf{R})|^2$ for all lattice vectors i.e. $u(\mathbf{r})$ must be periodic. Now if $|\psi(\mathbf{r})|$ has to be periodic then the only way $\psi(\mathbf{r})$ can differ from lattice point to lattice point is by a phase factor. It is possible to show that this phase factor has to be of the form $e^{i\mathbf{k}\cdot\mathbf{r}}$.

7.3 Energy-Bands

The Hamiltonian for an electron in a lattice consists of the sum of the kinetic energy term and the electrostatic potentials for the electron interacting with every atom in the lattice. i.e.

$$\hat{H}_{\text{crystal}} = -\frac{\hbar^2}{2m_e}\nabla^2 + \sum_{\mathbf{R}} U_{\text{atom}}(\mathbf{r} - \mathbf{R}) \tag{7.4}$$

where $\mathbf{R}$ is a lattice vector.

Now let us consider a simple cubic crystal made up of hydrogen atoms: we will restrict ourselves to considering the bands formed from the lowest energy s orbitals. The wavefunction we need therefore consists of a linear combination of the s orbitals on each of the atoms. We know that the wavefunction has to satisfy Bloch's theorem. One wavefunction that meets these requirements, and therefore the correct one, is:

$$\Psi(\mathbf{r}) = \frac{1}{\sqrt{N}}\sum_{\mathbf{R}}\Phi_S(\mathbf{r} - \mathbf{R})e^{i\mathbf{k}\cdot\mathbf{R}} \tag{7.5}$$

---

[1]See appendix A for a complete proof

where $\Phi_S(\mathbf{r})$ is the s orbital on the hydrogen atom at the origin and N is the number of atoms in the crystal. The periodic potential $U(\mathbf{r})$ due to each individual atom at site $\mathbf{R}$ can be written as the sum of each atomic potential U_{atom}:

$$U(\mathbf{r}) = \sum_{\mathbf{R}} U_{\text{atom}}(\mathbf{r} - \mathbf{R}). \tag{7.6}$$

This yields a time independent Schrödiger equation of the following form

$$
\begin{aligned}
E\,\Psi(\mathbf{r}) &= \left[-\frac{\hbar^2}{2m}\nabla^2 + U(\mathbf{r}) \right] \Psi(\mathbf{r}) \\
&= \frac{1}{\sqrt{N}} \sum_{\mathbf{R}'} \left[-E_{\text{atom}} + \sum_{\mathbf{R} \neq \mathbf{R}'} U_{\text{atom}}(\mathbf{r} - \mathbf{R}) \right] e^{i\mathbf{k}\cdot\mathbf{R}'} \Phi_S(\mathbf{r} - \mathbf{R}'),
\end{aligned}
\tag{7.7}
$$

where E_{atom} is the energy eigenvalue for a single isolated atomic orbital with the following form

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + U_{\text{atom}}(\mathbf{r} - \mathbf{R}) \right] \Phi_S(\mathbf{r} - \mathbf{R}) = -E_{\text{atom}}\,\Phi_S(\mathbf{r} - \mathbf{R}). \tag{7.8}$$

In order to determine the energies of the electrons in a crystal we now have to calculate:

$$
\begin{aligned}
\langle E \rangle &= \int \Psi^* \hat{H}_{\text{crystal}} \Psi \, d\mathbf{r} \\
&= \frac{1}{N} \int \sum_{\mathbf{R}''} \Phi_S^*(\mathbf{r} - \mathbf{R}'') e^{-i\mathbf{k}\cdot\mathbf{R}''} \hat{H}_{\text{crystal}}(\mathbf{r}) \sum_{\mathbf{R}'} \Phi_S(\mathbf{r} - \mathbf{R}') e^{i\mathbf{k}\cdot\mathbf{R}'} \, d\mathbf{r}
\end{aligned}
\tag{7.9}
$$

Substituting the right-hand-side of (7.7) and (7.4) into (7.9) then yields

$$\langle E \rangle = -E_{\text{atom}} + \frac{1}{N} \sum_{\mathbf{R}',\mathbf{R}''} e^{i\mathbf{k}\cdot(\mathbf{R}' - \mathbf{R}'')} \int \Phi_m^*(\mathbf{r} - \mathbf{R}'') \sum_{\mathbf{R} \neq \mathbf{R}'} U_{\text{atom}}(\mathbf{r} - \mathbf{R}) \Phi_S(\mathbf{r} - \mathbf{R}') \, d\mathbf{r}. \tag{7.10}$$

This expression looks very complicated however we can easily separate it into different types of terms. The only remaining non-zero terms occur when $\mathbf{R}'' = \mathbf{R}$ or $\mathbf{R}'' = \mathbf{R}'$. If we first consider the case when $\mathbf{R}'' = \mathbf{R}'$, the numerator in (7.10) becomes

$$-\frac{1}{N} \sum_{\mathbf{R}'} E_U$$

where

$$-E_U = \int \Phi_S^*(\mathbf{r} - \mathbf{R}') \sum_{\mathbf{R} \neq \mathbf{R}'} U_{\text{atom}}(\mathbf{r} - \mathbf{R}) \Phi_S(\mathbf{r} - \mathbf{R}') \, d\mathbf{r}. \tag{7.11}$$

This is the contribution to the expectation value for each ion due to the potential of all the neighbouring ions. This quantity will have the largest contribution to the energy expectation value. The remaining term occurs when $\mathbf{R}'' = \mathbf{R}$ and is given by

$$-\frac{1}{N} \sum_{\mathbf{R}', \mathbf{R} \neq \mathbf{R}'} e^{i\mathbf{k}\cdot(\mathbf{R}' - \mathbf{R})} E_{\text{next}}$$

where

$$-E_{\text{next}} = \int \Phi_S^*(\mathbf{r} - \mathbf{R}) \sum_{\mathbf{R} \neq \mathbf{R}'} U_{\text{atom}}(\mathbf{r} - \mathbf{R}) \Phi_S(\mathbf{r} - \mathbf{R}') \, d\mathbf{r}. \tag{7.12}$$

Equation 7.12 describes the contribution to the expectation value resulting from wavefuction overlap of nearest neighbour orbitals. The quantity E_{next} is obtained from the atomic wavefunction, potential and nearest neighbour distance. As the potential due to each ion is attractive, it is conventional to denote the quantities $E_{\text{atoms}} = E_{\text{atom}} + E_U$ and E_{next} with a preceding negative sign.

Terms where both orbitals in the above integral are on the same atom, e.g. $\Phi_S(\mathbf{r})\hat{H}_{\text{crystal}}(\mathbf{r})\Phi_S(\mathbf{r})$, have the largest contribution to the total energy, $-E_{\text{atoms}}$. The next biggest terms in the Hamiltonian are those involving an orbital on one atom and an orbital on one of its nearest neighbours, e.g. $\Phi_S^*(\mathbf{r} - \mathbf{a}_x)e^{i\mathbf{k}\cdot\mathbf{a}_x}\hat{H}_{\text{crystal}}(\mathbf{r})\Phi_S(\mathbf{r})$. The magnitude of $-E_{\text{next}}$ depends on how much of the orbital extends in the direction of the neighbour atom, i.e. how much it overlaps the next atom, and is therefore often called the overlap integral. For a pictorial example of this see Fig. 7.2.
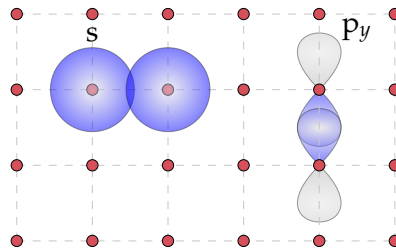

Figure 7.2: Schematic representation of overlap between s and p orbitals with simple cubic lattice. s-orbital has equal overlap with each of its nearest neighbours. The p-orbital has larger overlap with atom in direction in which it points. The overlap in this direction is greater than that of the s orbital.

For a simple cubic crystal, each atom is surrounded by six nearest neighbours and for s-orbitals, the overlap is the same for each. We can then write the overlap contribution to the energy expectation value as:

$$\frac{1}{N}\sum_{\mathbf{R}}e^{-i\mathbf{k}\cdot\mathbf{R}}E_{\text{next}} = 2E_{\text{next}}\left[\cos(k_xa_x) + \cos(k_ya_y) + \cos(k_za_z)\right] + ... \tag{7.13}$$

Additional cosine terms will appear as more distant neighbour ions are included in the summation and will provide additional corrections to the energy bands. The energy of the Bloch state is given by:

$$E_{\text{total}}(\mathbf{k}) = -E_{\text{atoms}} - 2E_{\text{next}}\left[\cos(k_xa_x) + \cos(k_ya_y) + \cos(k_za_z)\right] \tag{7.14}$$

The influence of the overlap integral on the energy band structure is immediately apparent as a results of evaluating the overlap term of a simple cubic crystal with a single atomic s-orbital. If the overlap wavefunction is small, the quantity E_{next} is also small resulting in a comparatively narrow band width given by $4E_{\text{next}}$. If the separation between ions is decreased, the overlap of neighbouring wavefunctions will increase as will the magnitude of the overlap integral and hence the width of the band. We therefore expect the atomic energy levels of free atoms to broaden in width as distance between atoms is decreased (Fig. 7.3). This broadening effect can be so large that the bands begin to overlap allowing electrons to migrate continuously from one band to the other. We might therefore conclude that the conductivity will increase continuously with a reduction in the interatomic spacing. A more detailed calculation that goes beyond the independent electron approximation will likely predict an abrupt increase in
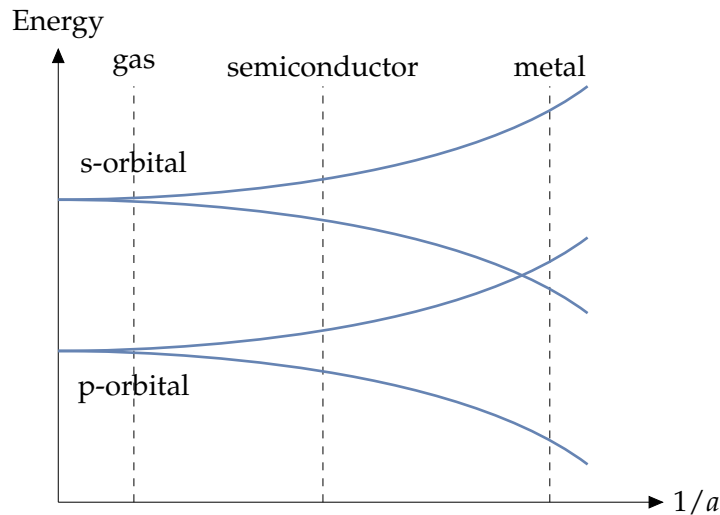
Energy

gas   semiconductor   metal

s-orbital

p-orbital

$1/a$

Figure 7.3: Schematic representation of overlap between s and p-orbitals with simple cubic lattice. The s-orbital has equal overlap with each of it's nearest neighbours. The p-orbital has larger overlap with atom in direction in which it points. The overlap in this direction is greater than that of the s-orbital.

the conductivity at a specific interatomic distance as the ions are brought into close proximity. This is due to the initial repulsive force that an electron experiences at an atomic site that is occupied by another electron. The transition occurs when this repulsive force is overcome by the attraction of the ions and the material transitions from insulating to metallic (*Mott transition*).

7.4   Crystal Momentum

So far we have not discussed whether **k** has a physical meaning or whether it is just a mathematical number. Mathematically **k** is just a way of quantifying how quickly the phase part of the Bloch function varies in space. Physically it turns out that $\hbar$**k** has many of the properties of momentum and is therefore called the crystal momentum of the electron. In general however it is not accompanied by any real momentum of the crystal as a whole. As a crystal only possess the translational symmetry of its lattice, the resulting momentum conservation law is only conserved upto an additive reciprocal lattice vector.[2] For instance if two electrons collide then we can write down an equation for the conservation of crystal momentum,

$$\mathbf{k}^1_{\text{in}} + \mathbf{k}^2_{\text{in}} = \mathbf{k}^1_{\text{out}} + \mathbf{k}^2_{\text{out}} + \mathbf{G} \tag{7.15}$$

where **G** is any reciprocal lattice vector. In addition when we apply an electric field to a crystal the electron's crystal momentum increases in accordance with:

$$\hbar \frac{d\mathbf{k}}{dt} = -e\mathbf{E} \tag{7.16}$$

---

[2]Translational invariance of free space gives rise to the law of momentum conservation.

However in a strict sense the crystal momentum is not the momentum of the electron, which is flying around in the atomic orbitals, but is instead more like the average momentum of the electron.

7.5   Dispersion Relation

Thus $E_{\text{total}}(\mathbf{k})$ is the tight binding dispersion relation. From it and Equ. 6.27 we can calculate the tight binding effective mass for a cubic crystal as:

$$m_e^* = \frac{\hbar^2}{2a^2 E_{\text{next}} \cos(ka)}.$$

Thus as the overlap, E_{next}, increases the mass decreases.

7.6   Relationship between Tight Binding and Nearly Free Electron Models

In Chapter 5 you were introduced to another model for how a crystal effects the energy of electrons within it called the nearly free electron model. This starts by assuming that the electrons are free and that the crystal weakly perturbs the electrons leading to energy gaps opening up in the dispersion relation. The Tight Binding model starts by assuming that the electrons are localised on atoms and that the crystal weakly perturbs this by allowing electrons to move from atom to atom. In the free electron model we get a change-over in behaviour from a metal to a semiconductor to an insulator as the energy gaps increase in magnitude[3]. In the Tight Binding model we start with gaps between the atomic levels and we get a changeover from insulator to semiconductor to metal as the energy width of bands increases bringing them closer until they overlap in energy leading to a metal (see footnote and Fig. 7.3). Now from all these differences you might think that the two models are completely different however they have very important similarities. In particular they are both single electron models and most importantly they lead to the same predictions for the behaviour of electron energy levels in crystals.

7.7   Interpretation of the Bandstructure of Semiconductors using the Tight Binding Model.

One of the most useful aspects of the Tight Binding model is that it provides us with a simple way of looking at the nature of the energy bands in a solid. Each band can be associated with a single or group of atomic orbitals. For instance in the case of GaAs and other III-IV semiconductors it turns out that the valence bands have a p-orbital-like character and the conduction band has an s-orbital like character. Now optical transitions in solids are controlled by the same selection rules as in atoms i.e. we can only increase the orbital angular momentum by one $\Delta j = \pm 1$. This would mean that if instead of a p character the valence band of GaAs had a d orbital character then optical transitions between the valence and conduction band would not be allowed which would mean no red LED's or CD players.

---

[3]This statement assumes that the atomic orbitals/bands are full in the case of large atomic separation.

Chapter 8

Magnetic Properties of Crystals

Magnetism in solids is a topic of both great fundamental interest and also technological importance. When people think of magnetic technologies they tend to focus on data storage in hard drives and magnetic tape, however long before these applications of magnetism the generation, distribution and application of electrical energy required a good understanding of magnetism in solids. In the future people have suggested that all magnetic computers might take over from current technology, so called spintronics, and that it might even be possible to develop *quantum computers* based on the spin (magnetic) moment of individual electrons or atoms.

Like most of the topics covered in this course, magnetism is far too vast a subject to cover all the details and so in the course we will instead introduce some of the basic experimental phenomena and models of the field. I will assume you understand the basic concepts from electromagnetism of the definition of the magnetisation, **M** and magnetic field, **B**. We will also need the magnetic susceptibility χ, which is defined by $\mu_0 \mathbf{M} = \chi \mathbf{B}$. I will also assume you are familiar with Hund's rules from Atomic Physics (PHYS3008). The level of detail required for this course is close to but slightly less than that presented in Myers. If you are interested in a more fundamental and quantum mechanical based approach to magnetism you might like to look at Ashcroft and Mermin.

8.1 Diamagnetism

Diamagnetism is the term used to describe the magnetic behavior where an applied magnetic field acts to produce magnetisation in a material to oppose the applied field whose magnitude is proportional to the field strength (see Fig. 8.1). That is the magnetic susceptibility is negative.

Diamagnetism comes in many different forms, including perfect diamagnetism, i.e. all magnetic field is excluded from the material, associated with superconductivity. Due to the fact that all atoms consist of electrons orbiting nuclei all materials have a diamagnetic component to their magnetic response, called Langevin diamagnetism. Although it is always present it has a very small magnitude and is often swamped by other magnetic responses. We will now look at a fairly simple classical model which allows us to understand where this diamagnetic response comes from and which can be used to make a good prediction of the magnitude.
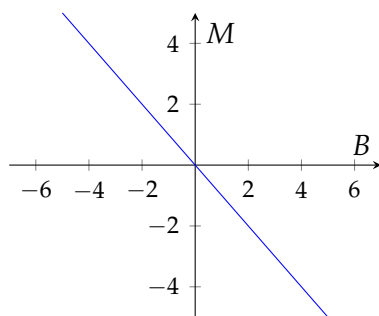
Figure 8.1: Magnetisation (*M*) versus Magnetic Field (*B*) for a Diamagnetic Substance.

The model starts by assuming that the electrons orbiting around the nucleus of an atom can be considered to be like electrons flowing in a loop of wire whose area is the same as the average area of the orbital in which the electrons reside. When a magnetic field is applied to such a system we know from Lenz's law that an e.m.f. will be generated in such a loop whose direction is such as to generate a current in the loop which will act to produce a magnetic field which will oppose the change in the magnetic field, i.e. a diamagnetic response. A more detailed, mathematical formulation of this model predicts that the Larmor susceptibility is given by

$$\chi = -\frac{e^2 \mu_0}{6m} \frac{N}{V} \sum_{i=1}^{Z} \langle R_i^2 \rangle \tag{8.1}$$

where N/V is the number density of unit cells in the material and the sum is the sum of the average area, perpendicular to the applied magnetic field, of all the occupied electronic orbitals of the atoms making up the unit cell of the material. As can be seen from this expression the Langevin diamagnetism is independent of temperature.

8.2   Localised versus Delocalised Electrons

In the discussion of the Langevin diamagnetism we implicitly assumed that the electrons in a material are in atomic like orbitals around a particular atomic nuclei, i.e. that the electrons are localised on a single atom. We know this is not always true. In particular in metals we have nearly free electrons. However in many materials most of the electrons, the core electrons, are in atomic like orbitals and in some materials such as ionic insulators, e.g. NaCl, all the electrons are localised in atomic orbitals. In fact although delocalised electrons do not show Langevin diamagnetism they show an equivalent, although larger in magnitude, diamagnetic response called Landau diamagnetism. In general as the magnetic properties of materials depends on the nature of the electronic states the differences between localised and delocalised electrons will lead to differences in the magnetic response of metals and insulators.

8.3   Paramagnetism

Paramagnetism is the name for the magnetic phenomenon where already existing magnetic moments, from either atoms/ions with permanent magnetic moments or free electrons, align parallel to the magnetic field. This leads to a magnetic response that is linear in magnetic field

at low magnetic fields with a positive susceptibility. At high enough fields the magnetisation saturates as all the magnetic moments become fully aligned (Fig. 8.2).

To start the discussion on paramagnetism we will discuss the paramagnetic behaviour of a gas of non-interacting atoms/ions that have a permanent magnetic dipole. Although you might think this would never apply to the case of crystalline solids actually for some solids this model applies very well with only a few simple modifications.

From Atomic Physics we know that the total angular momentum of an atom/ion is a good quantum number of its ground state. For any particular angular momentum J there exist $2J + 1$ different states that are characterised by the quantum number m_J, which is the component of the angular momentum in a specified direction (which is referred to as the z axis) and can take the values J-integer from J to $-J$. These states are degenerate if the atom is not in a magnetic field. Associated with the total angular momentum is a permanent magnetic moment, $g\mu_B J$, where μ_B is the Bohr magneton and g is the Landé splitting factor given by:

$$g = 1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)} \tag{8.2}$$

where S and L are the total spin and orbital angular momentum quantum numbers of the ground state of the atom. The magnetic moment in any particular direction, designated as the z direction, is given by $-g\mu_B m_J$. In the presence of a magnetic field, which defines the z-axis direction, the atoms/ions magnetic dipole has potential energy given by $-\mathbf{m} \cdot \mathbf{B}$, where $\mathbf{m}$ is the magnetic moment. We will assume that $\mathbf{m}||\mathbf{B}$ and hence this energy becomes: $-mB$. States with different z components of angular momentum, m_J, differ in energy from their original energy by $g\mu_B m_J B$. This splitting in energy means that the lower energy states are more likely to be populated than the higher energy states. Using statistical mechanics (PHYS2024) we can write down that the mean magnetic moment of the atom/ion is given by:

$$m_{\text{avg}} = \frac{\sum_{m_J=-J}^{J} g\mu_B m_J \, e^{-g\mu_B m_J B / k_B T}}{\sum_{m_J=-J}^{J} e^{-g\mu_B m_J B / k_B T}} \tag{8.3}$$

Which, using the rules for geometric series, can be shown to be[1]:

$$m_{\text{avg}} = g\mu_B J \left[\frac{2J+1}{2J} \coth\left(\frac{\alpha(2J+1)}{2J} \right) - \frac{1}{2J} \coth\left(\frac{\alpha}{2J} \right) \right] \tag{8.4}$$

$$\alpha = \frac{g\mu_B J B}{k_B T} \tag{8.5}$$

This expression for m_{avg} divided by the prefactor $g\mu_B J$ is called the Brillouin function, $B_J(\alpha)$. The dependence of this function on α for a range of different J are shown in Fig. 8.2. The Brillouin function has the dependence on magnetic field we were expecting with saturation at high fields (α) and a linear dependence of magnetisation on field at low fields. In the limit of small fields the above expression gives the magnetic susceptibility to be:

$$\chi = \frac{Ng^2\mu_B^2\mu_0}{3k_B T} J(J+1) = \frac{C}{T} \tag{8.6}$$

---

[1]See appendix B

The second form of the expression is an expression of Curie's law. Curie's law, which only applies in the limit of small fields, was "discovered" from experimental data long before the explanation of it in terms of quantum mechanics was made.
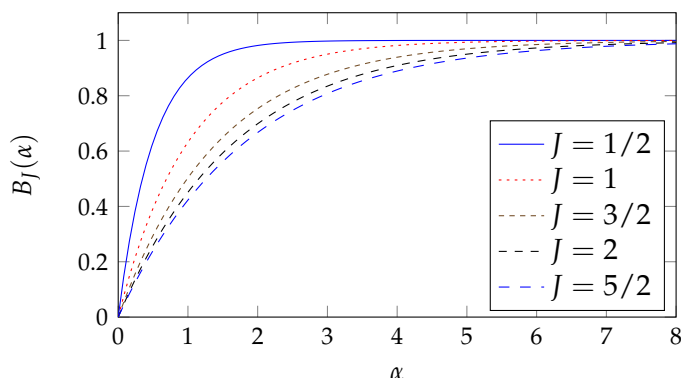

Figure 8.2: The Brillouin function for the cases $J = 1/2$ to $5/2$.

The reason why we started discussing this model for paramagnetism is because we wanted to apply it to solids. Under what conditions would we expect this model to apply? Well we assumed that the magnetic atoms were independent of each other i.e. that the interactions between atoms are weak on the energy scale of interest. Thus we would expect the model to work best at higher temperatures and magnetic fields in solids with weak interactions between magnetic atoms. This is the case in the so called dilute magnetic salts in which a small number of magnetic ions are diluted by a large number of other atoms/ions, e.g. $MnSO_4(NH_4)_2SO_4.6H_2O$. And in fact the theory works extremely well in such compounds containing magnetic ions with magnetic moments due to unpaired electrons in f-orbitals. However although it is possible to fit the magnetic field and temperature dependence of the magnetisation of such compounds with ions with partially filled d-orbitals the values of $C/(\mu_0\mu_B^2/3k_B)$, which should equal $g^2J(J+1)$, obtained from such fits are not in agreement with those calculated by Hund's rules (see Table 8.1). Much better agreement to the experimental values is obtained if it is assume that L for these atoms is actually zero. This effect is known as quenching of the orbital angular momentum.

Table 8.1: Comparison of measured Curie constants per atom, scaled by the prefactor in the equation above, to values predicted by theory with and without orbital quenching for examples of ions with partially filled d and f bands.

Ion	Configuration	Ground State (in gas)	$g^2J(J+1)$	$4S(S+1)$	$C/(\mu_0\mu_B^2/3k_B)$
Pr^{3+}	$4f^2 5s^2 p^6$	$^3H_4$	12.8	8	12.3
Er^{3+}	$4f^{11} 5s^2 p^6$	$^4I_{15/2}$	92	24	88.4
Dy^{3+}	$4f^9 5s^2 p^6$	$^6H_{15/2}$	113	120	112
Ce^{3+}	$4f^1 5s^2 p^6$	$^2F_{5/2}$	6.5	8	5.76
Mn^{3+}	$3d^4$	$^5D_0$	0	48	48.0
Cu^{2+}	$3d^9$	$^2D_{5/2}$	12.6	3	3.6
Cr^{3+}	$3d^3$	$^4F_{3/2}$	0.6	3	3.6

The reason that the orbital angular momentum is quenched is the effect of the surrounding atoms on the energy of the d-orbitals leading to a breakdown in the degeneracy of the different

m_L state; the so called crystal field splitting of the orbital angular momentum states. To see how this might be consider an ion with d-orbitals in close proximity to four oppositely charged ions. The electrons on the "magnetic" ion experience the electric field due to the other ions. This field means that unlike the case where the magnetic ion is in a gas, the potential the electrons experience only has discrete rotational symmetry. This means that the orbital angular momentum states are no longer good quantum numbers instead these states are combined to form new states where the electrons on the magnetic ion are either localised near, lower energy, or far from the surrounding atoms. The new states will not have a sense of rotation associated with them and therefore do not have a magnetic moment. The magnitude of the crystal field effect depends on how strongly the electrons in the orbitals experience the electric field from neighbouring atoms. The d-orbitals stick out a long way from the atom bringing the electrons in them near to neighbouring atoms. They thus have a large crystal field splitting, and are quenched. Where as f-orbitals, which are shielded by electron in s-orbitals and d-orbitals from the electric field of the surrounding atoms, are not quenched.

8.4   Pauli Paramagnetism

So far we have only discussed paramagnetism for insulating solids. Metals are also paramagnetic; they show so called Pauli paramagnetism. Pauli paramagnetism comes about because the conduction electrons have spin. Any applied magnetic field causes the energy of electrons with the same wavevector to depend on the direction of their spin, i.e. electron with spin/magnetic moments in the direction of the magnetic field go down in energy by $\mu_B B$ and those opposite in direction will go up in energy by the same amount. One way of describing this is to draw out separate bands for up and down spins (see Fig. 8.3).
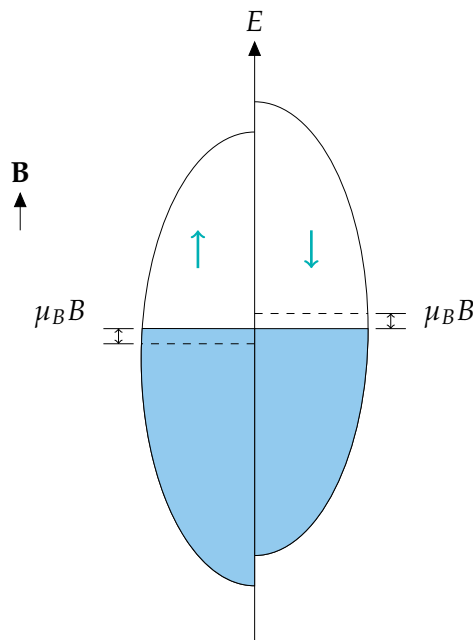

Figure 8.3: Effect of magnetic field on the two different, parallel and anti-parallel, spin bands of a metal. The dashed line shows the position of the Fermi energy in each band before the application of the magnetic field, the shaded states are filled. The arrows show the direction of electron spin, relative to the magnetic field B.

As electrons of one spin can scatter into states with opposite spin the populations of the electrons with spins parallel and anti-parallel to the magnetic field will have the same chemical potential (Fermi energy). Thus the total number of electrons with spin parallel will increase and with spin antiparallel will decrease. The imbalance in the two populations is given by the difference in energy for spin parallel and anti-parallel times the density of states at the Fermi energy for each band $\frac{1}{2}D(E_F)$, i.e. half the density of states for a band with both spins, which gives a difference in population of $\mu_B D(E_F)B$. The imbalance in population will lead to the electrons having a net magnetic moment for each spin of μ_B. This leads to a susceptibility of $\mu_0 \mu_B^2 D(E_F)$.

Note that the above discussion simplifies the magnetic response of a metal a great deal. In particular as we have already said metals also have a diamagnetic response, Landau diamagnetism. This comes from the orbital part of the electron wavefunction.

8.5   Ferromagnetism

Ferromagnetism is the property that at temperatures below some critical value a material has a permanent magnetic moment even in the absence of a magnetic field[2]. Above this critical temperature materials that display ferromagnetism behave as paramagnets although with a susceptibility whose temperature dependence does not follow Curie's law (see Fig. 8.4).
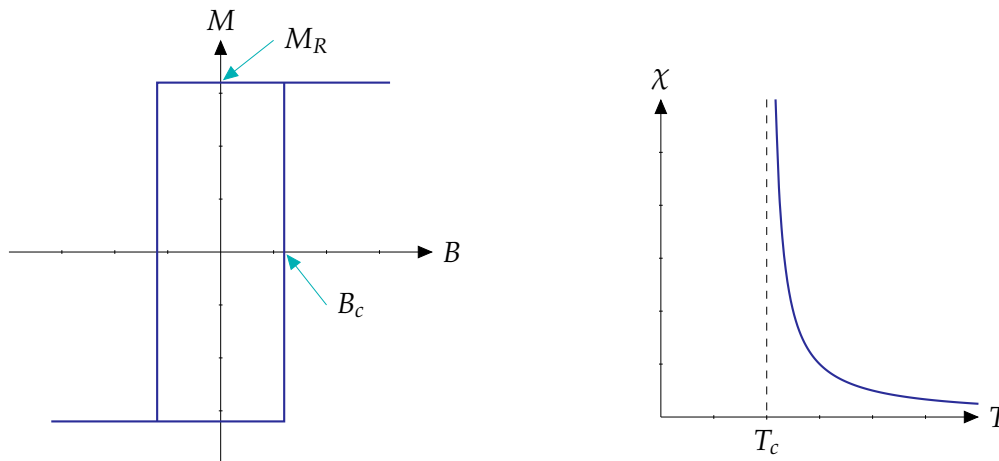

Figure 8.4: (Left) Magnetisation, M, versus magnetic field, B, for a ferromagnet for $T < T_c$. M_R is the remanent magnetisation and B_C is the coercive field. (Right) Schematic temperature dependence of the magnetic susceptibility of a ferromagnet. χ tends to infinity as T tends to T_c and is not defined for T less than T_c.

As having a permanent magnetic moment at zero magnetic field implies that any particular moment in the material must know which way the other magnetic moments in the material are pointing, ferromagnetism implies that their must be some interaction between the magnetic moments in the material. One obvious possibility is that the magnetic field from one moment aligns its neighbours. However it is clear from fairly simple calculations, that the magnetic interactions between moments is several orders of magnitude too small to be responsible for ferromagnetism in most materials. Instead the moments interact indirectly via the exchange

---

[2]Although microscopically this is always true the existence of magnetic domains means that not all objects made of ferromagnetic material will behave as permanent magnets for $T < T_c$. See discussion of magnetic domains.

interaction. The exchange interaction comes about because the coulomb energy due to two electrons depends on their spin.[3] Electrons with parallel spins have to have anti-symmetric spatial wavefunctions; this keeps the two electrons apart leading to a decrease in the energy due to the coulomb repulsion between the two electrons. It is the exchange interaction that leads to the Hund's rule that S should be maximised. The details of the exchange interaction varies in different materials and are complicated. However we can go a long way towards understanding ferromagnets with the fairly simple model that the effect of the exchange interaction between the magnetic moments can be treated as if any particular magnetic moment in a material was paramagnetic, i.e.

$$M = g\mu_B NJ \left[\frac{2J+1}{2J} \coth\left(\frac{\alpha(2J+1)}{2J} \right) - \frac{1}{2J} \coth\left(\frac{\alpha}{2J} \right) \right] = g\mu_B NJB_J(\alpha) \tag{8.7}$$

where N is the number density of magnetic ions in the material, and $\alpha = \frac{g\mu_B JB_{\text{local}}}{k_B T}$.

And simultaneously each moment experienced the effect of all the other moments in the material as a component of the local magnetic field whose magnitude was proportional to the average magnetisation of the material, M:

$$B_{\text{local}} = B_{\text{applied}} + \lambda\mu_0 M \tag{8.8}$$

where λ parametrises the strength of the exchange interaction between magnetic moments.

The first question we should ask is does this model lead to the sudden appearance of a spontaneous, i.e. $B_{\text{applied}} = 0$, magnetisation with temperature. To answer this we need to see whether there are non-zero solutions for M of the equation:

$$M = g\mu_B NJB_J \left(\frac{g\mu_B J\lambda\mu_0 M}{k_B T} \right) \tag{8.9}$$

It is not possible to solve the above equations analytically however it is possible to see when a solution exists using a graphical technique as shown in Fig. 8.5.

In the figure the two equations:

$M = g\mu_B NJB_J(\alpha)$

$M = \frac{B_{\text{local}}}{\lambda\mu_0} = \frac{K_B T}{g\mu_B J\lambda\mu_0}\alpha$

are plotted on a graph of $M/g\mu_B JN$ versus α. Any point where the lines representing these two equations touch is a simultaneous solution of both equations. At high temperatures the straight line representing equation 2 has a very high gradient and the two lines only intercept at $M = 0$, i.e. no spontaneous magnetisation. However, at some point as the temperatures decrease the gradient of the straight line will be such that the lines cross at finite M. This will occur when the gradient of the first equation near $M = 0$, i.e. the susceptibility of the paramagnet at small fields $\chi = C/T$, equals $1/\lambda$, i.e. the critical temperature $T_c = C\lambda$. At temperatures beneath this critical temperature there exists three simultaneous solutions to the above equation. Two with finite, equal magnitude but opposite sign values of M and one with M equals zero. In order to observe a spontaneous magnetisation the zero solution has to be unstable and the other two solutions have to be stable. A stable solution being one for which

---

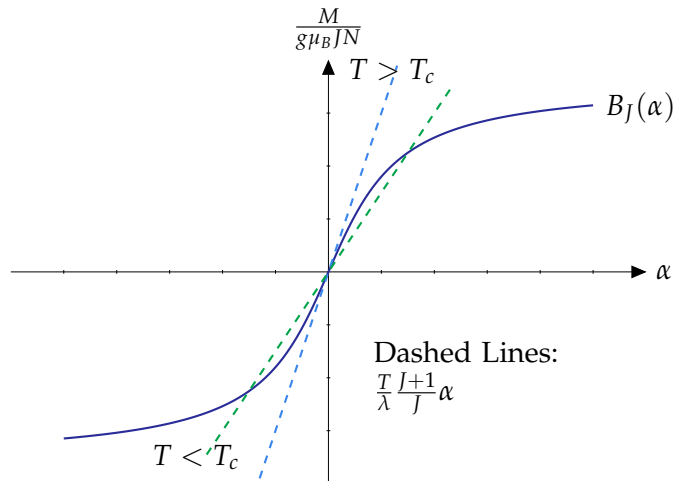[3]See Appendix C for a discussion of the exchange interaction.

Figure 8.5: Graphical solution of model for ferromagnetism described in the text for two different temperatures T_1 and T_2 either side of the critical temperature T_c.

if the magnetisation is perturbed from the solution it will return back to the original value. Consider the situation that M is initially zero and is slightly perturbed positively ($M = M_1$). This will lead to a local magnetic field for which we can determine, $\alpha = \alpha_1$, from the graph, Fig. 8.6, by drawing the line $M = M_1$ and finding where it intercepts the line representing the second equation. This local field will induce magnetisation, M_2, which we can determine by drawing the line $\alpha = \alpha_1$ and finding where it intercepts the line representing the first equation. It should already be clear that if we repeat this process that the magnetisation will continue to grow and $M = 0$ is an unstable solution.
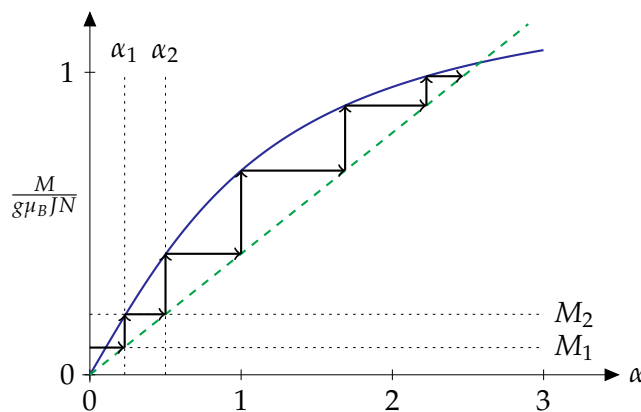

Figure 8.6: Graphical proof that for low temperatures the $M = 0$ solution is unstable. M_1 and α_1 are described in the text and the arrows are present to show path between values of M and α.

If we do the same analysis at the finite solutions we would find they were both stable. So we conclude that the model does give the correct ferromagnetic behaviour.

What about the magnetic field dependence of the magnetisation of a ferromagnet. For $T > T_c$ we know that for small magnetic fields any magnetisation will be small thus the equation for the magnetisation of a paramagnet can be approximated by Curie's law. Thus, $\mu_0 M = \frac{C}{T}(B_{\text{applied}} + \lambda \mu_0 M)$ or $\mu_0 M = \frac{C}{T - T_c} B_{\text{applied}}$ where the second expression is a statement of the Curie-Weiss law (Fig. 8.4).

For $T < T_c$ we have a permanent magnetic moment. What happens when we apply a magnetic field to a ferromagnet in this state. Well let us assume that the magnetic moment in the material can point in any direction, an assumption that is often false i.e. the magnet is isotropic, then the most important effect of an applied magnetic field will be to line up the magnetic moment of the material with the applied field. This leads to a magnetisation versus magnetic field of the form shown in Fig. 8.7. The vertical line at $B = 0$ is what we would expect for a material with infinite magnetic susceptibility.
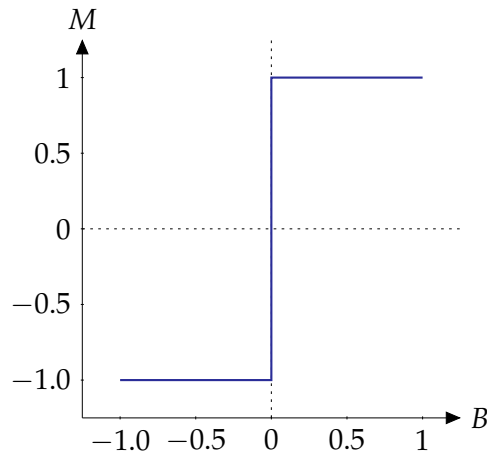
Figure 8.7: Magnetisation versus magnetic field for an isotropic ferromagnet.

If instead we assume that we are dealing with a ferromagnet where the magnetic moment may only point either direction along one line, i.e. infinite anisotropy. Then we have to use the full expression for the magnetisation of a paramagnet and again we have to solve either graphically or computationally. The effect of the applied magnetic field on the graphical solution is that the line for equation 8.7 no longer passes through the origin but has an offset along the α axis. Consider the situation set out in Fig. 8.8 where initially $B_{\text{applied}} = 0$ and M is positive. A magnetic field is then applied in opposition to the magnetisation with increasing magnitude. Initially the magnetisation slowly decreases until some critical field, the coercive field when the positive M solution no longer exists and the only solution is a large negative field at which point the magnetisation direction will suddenly reverse. If the magnetic fields magnitude was then decreased the magnetisation would remain negative i.e. the model shows the hysteresis expected from a ferromagnet.

Although this behaviour does indeed agree qualitatively with experiments, in fact the experimental behaviour depends on a number of factors not taken into consideration by this model. These include, in addition to the effect of finite anisotropy that will lower the coercive field, the demagnetisation field, analogous to the depolarisation field in the case of dielectric properties, which depends on the geometry of the sample, and the existence of magnetic domains.

The concept of magnetic domains, see Fig. 8.9, also explains why although the model we have just discussed predicts that all lumps of iron should be powerful magnets, as the critical temperature for iron is far above room temperature, we know this is not true. The basic concept is that a bulk piece of ferromagnet is made up of many different regions called domains within which all the magnetic moments are aligned as expected for a ferromagnet however in general the moment of different domains points in different directions leading to zero net magnetic moment for the bulk piece. From the theory we discussed above we know that having regions with opposite magnetic moment nearby must cost energy so why do domains
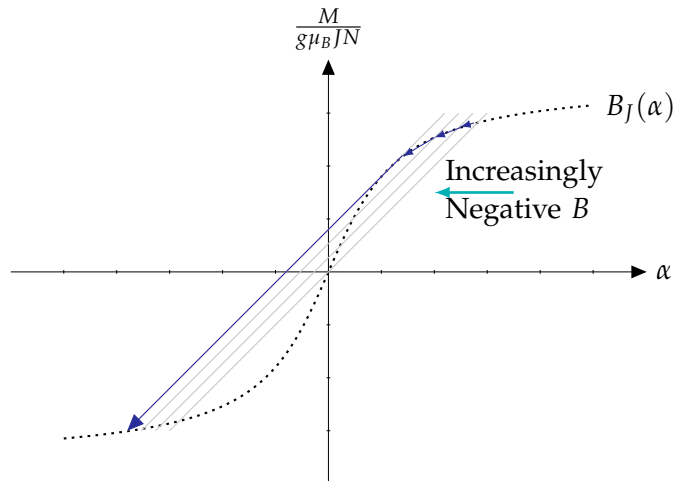
Figure 8.8: Graphical solution of the magnetisation of a ferromagnet as an increasing magnetic field is applied in opposition to the spontaneous magnetisation. The arrows show the path of the magnetisation with increasing magnetic field.
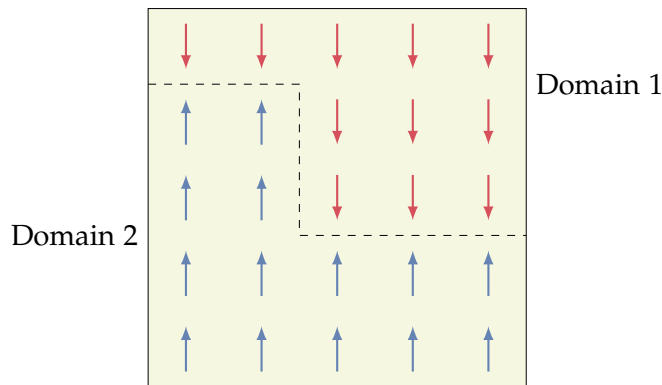

Figure 8.9: Schematic representation of a ferromagnetic particle with two domains.

occur. The main reason comes from the fact that a sample without domains will produce a magnetic field outside the material. This magnetic field has energy associated with it, proportional to B^2 integrated over all space (see electromagnetism notes). The energetic cost of the field will be proportional to the total magnetic moment of the magnetic object, which scales with the volume of the material. However the energy cost for forming domains is due to the surface area of the region between two domains. Therefore for big enough samples the formation of domains is energetically favourable. The critical size is very small in most cases, being approximately 100nm, and so nearly all magnetic material contains domains. When a magnetic object with domains is placed in a magnetic field the object can become magnetised either by the flipping of the magnetic moment in individual domains, less common, or by the movement of the boundaries of domains so that domains whose moment is parallel to the applied field grown in size and those whose moment is anti-parallel decrease in size. As there would at first appear to be no energy cost to moving the domain wall boundary apart from the magnetic field energy which varies continuously with magnetisation one might think that the magnetisation of a ferromagnet would be proportional to the applied field, i.e. a paramagnet. This is not what is observed in general. This is because the domain boundaries have lower energy when they can coincide with a defect within the material like a grain boundary in polycrystalline material. This effect is referred to as pining of the domain wall bound-

ary. The energy barrier to the change of domain size is what leads to magnetic hysteresis in bulk ferromagnets. As this energy cost depends on the microstructure of the material it can be manipulated by making/treating the material in different ways and so the behaviour of ferromagnetic materials is sample dependent. In the case that the energy barrier is high the material can sustain a higher applied magnetic field without the domains in it moving, i.e. it has a high coercive field.

8.6   Antiferromagnets and Ferrimagnets

So far we have talked about materials that have only one magnetic atom/ion in them. Of course magnetic materials exist containing two or more chemically different ions in them and/or two atoms/ions of the same element with different neighboring atoms. In this case a number of different arrangements of magnetic moments could be possible. It is possible in this case that the interactions between magnetic moments can lead to ordered low temperature states in which neighboring moments are anti-parallel. If the opposing magnetic moments are the same magnitude then the behavior is called antiferromagnetism. If the opposing magnetic moments are different magnitude the behavior is called ferrimagnetism.
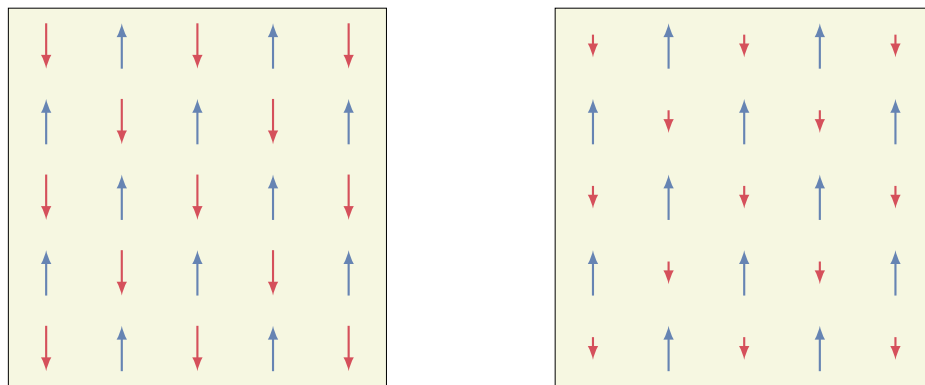

Figure 8.10: Schematic representation of an antiferromagnet (left) and ferrimagnet (right). The arrows indicate the direction and magnitude of the atomic magnetic moments.

8.7   Magnetic Data Storage

The availability of computing power has increased our appetite to handle large quantities of information, or ever more complex and storage-hungry types, such as pictures, soundtracks, and movies. All these have to be permanently stored and quickly retrieved, and this has spurred enormous developments in magnetic and optical storage technologies.

Familiar as you are with the concept of a bar magnet, their application to the storage of data using the direction that North points in, is nontrivial. Currents in a solenoid produce a magnetic field, and in the same way permanent magnets are produced by the angular momentum of charges. For instance in iron, more of the electrons near the Fermi energy are spinning one way than another - they like to line up together. How do we know the spins all line up? Once again we are faced with looking at microscopic regularly-spaced structures - in this case electrons on particular lattice sites whose spins point in a particular direction. And once again the solution is to go back to a diffraction technique, but we need one sensitive to the electron

spin rather than the electron charge. That is why neutron diffraction is so useful because it allows us to distinguish different spin configurations.

We will not go into why the spins want to align in this way in this course except to note that radically different magnetic properties can be controlled by constructing material structures which are small compared to the magnetic wavelengths - down to a few angstroms. The key element for magnetic storage is how small each bit can be made.

For instance the drive to create denser and denser material storage media means that each bit of information needs to be written in a smaller area (currently there are no volume methods commercially available for storing information). Magnetic data storage films consist of lots of small magnetic particles whose magnetic orientation is changed by a microscopic writing solenoid which passes overhead.

The problem of understanding how tiny magnetic moments can be flipped is an area of active research, because transiently each spin has to flip. They can do this individually or collectively. One physics problem facing the magnetic storage industry is how to deal with the fact that smaller and smaller magnetic bits start to have a finite probability for spontaneously flipping.

One other area of dramatic new research is to make miniature sensors to read the state of each magnetic bit as fast and reliably as possible. So a physical effect which depends on magnetic field is needed, and we have already looked at one of these - the Hall effect. Variants of the Hall effect where a B-field changes the device resistance have emerged which depend on how electrons with spin can tunnel from one magnet to another.
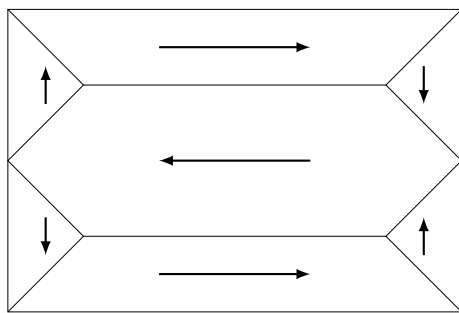

Figure 8.11: Illustration of magnetic domains at the surface of a crystal with cubic structure (e.g., SiFe). This configuration eliminates free poles at the ends of the sample by the introduction of triangular 'closure' domains.

Chapter 9

Semiconductors

9.1 Electron Occupation

Many semiconductors are now grown as large crystals for use in electronic devices. The ones we consider here are Si ($E_g \sim 1.15\text{eV}$), Ge ($E_g \sim 0.65\text{eV}$) and GaAs ($E_g \sim 1.52\text{eV}$). You should compare these bandgaps with the typical kinetic energy of free electrons at the first Brillouin zone edge $\sim 25\text{eV}$.

9.1.1 Intrinsic Semiconductors

To make a semiconductor conduct, we have seen that there needs to be a number of electrons in the higher-energy conduction-band states (or holes in the valence band). For small bandgaps, the energy available to electrons from thermal vibrations may be enough to excite some electrons from the valence band to the conduction band. The probability is given by:

$$P(e \text{ in conduction band}) \propto \exp\left(\frac{-E_g}{k_B T}\right) \tag{9.1}$$

Typically at room temperature $E_g/k_B T = 40$ so that there is little thermal excitation of electrons. As we cool these materials down the few excited electrons will fall back to the lower energy states, and become completely insulating (contrast this temperature dependence with that of metals). With an electron concentration n, and a hole concentration p, the rate of recombination must be proportional to np, since an electron and hole are annihilated together.

In equilibrium, the rate of generation must be the same as the rate of recombination. So we can write:

$$np = C \exp\left(\frac{-E_g}{k_B T}\right) \tag{9.2}$$

where C is a constant. In a pure semiconductor which starts with a completely filled valence band (this is called an *intrinsic* semiconductor) the electron and hole populations are always equal. So:

$$n_i = p_i = n_0 \, \exp\left(\frac{-E_g}{2k_B T}\right) \tag{9.3}$$

It is possible to calculate n_0, however for now let us consider the size of the intrinsic electron density in Si at room temperature:

$$n = n_i = 1.5 \times 10^{16} \text{ m}^{-3} \tag{9.4}$$

Hence roughly one electron for every 4μm x 4μm x 4μm cube, which is very tenuous!

The rate of generation which balances the recombination will not change even if the density of electrons or holes is modified, which means that the following holds even for non-intrinsic semiconductor materials:

$$np = n_i^2 \text{ (ALWAYS!)} \tag{9.5}$$

There is no rigid distinction between semiconductors and insulators. Often they are characterised according to their conductivities: semiconductors have $\sigma \sim 10^5$ to 10^{-5} Ω^{-1} m^{-1}, while insulators have lower σ down to 10^{-24} Ω^{-1} m^{-1}. Metals typically have $\sigma \sim 10^7$ to 10^8 Ω^{-1} m^{-1}.

Table 9.1: Real resistivities at 300 K.

Material Type	Material	Typical $\rho/\Omega m$
Superconductor		$0 \, (< 10^{-25})$
Good Metal	Ag	1.6×10^{-8}
	Cu	1.7×10^{-8}
	Au	2.2×10^{-8}
	Bi	1.2×10^{-6}
Alloy	constantin	4.4×10^{-8}
Semiconductor	Si	2×10^3
Insulator	fused quartz	$> 5 \times 10^{16}$
	teflon, polythene	$> 10^{20}$

Although the metals conduct very well at dc, remember that the resistance increases at higher frequencies. This is the reason that the new generations of 1GHz microprocessors will use Cu tracks, to decrease both heating losses, and radiation coupling between different wires on the chips.

9.1.2 Extrinsic Semiconductors

If we want to make semiconductors useful we have to find a way adding electrons so that a controllable number sit in the bottom of the conduction band where they can carry a current. Or equivalently we take away a few electrons so that a reasonable density of holes sits at the top of the valence band. By adding small concentrations of impurities or *dopants* to the semiconductor crystal, the electron and hole concentrations can be strongly modified. This produces an extrinsic semiconductor with $n \neq p$. If the dopant ion releases an extra electron (or hole), then the semiconductor is n-type (or p-type). Remember though that still the generation rate remains the same so that $np = n_i^2$.
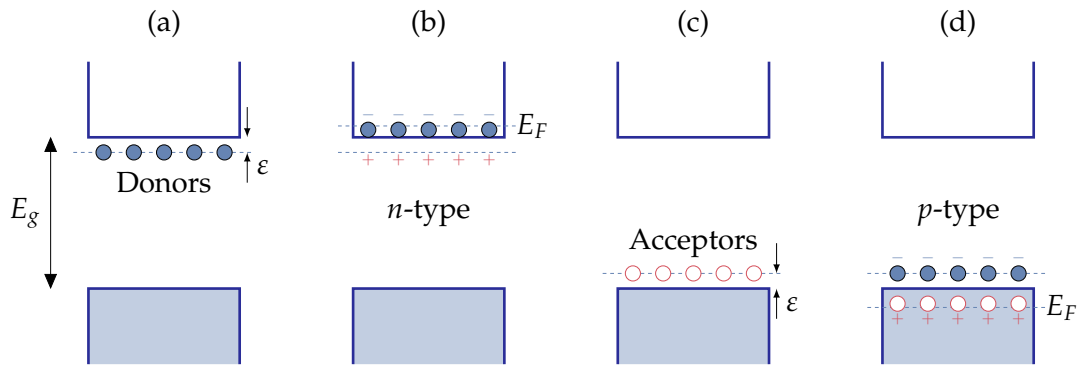
Figure 9.1: Doping of electrons ((a) - low Temperature, (b) - $T = 300$K) and holes (c)- low Temp, (d) - $T = 300$K).

9.2 Charge Transport

Electrical current is carried by both electrons and holes moving in opposite directions (since they have opposite charges) when an electric field is applied. This current is called the drift current and the model of jellium (Chapter 5) still works if we understand that there are two components, j_e and j_h. The electron and hole conductivities are generally different

$$\sigma_e = \frac{ne^2\tau_e}{m_e^*} \ , \ \sigma_h = \frac{pe^2\tau_h}{m_h^*} \tag{9.6}$$

since the electrons and holes have different densities, masses, and scattering times.

In an electric field the drift current can be written:

$$j_e = e \left(\frac{e\tau_e}{m_e^*}\right) n\mathcal{E} = e\mu_e n\mathcal{E} \tag{9.7}$$

where the electron mobility $\mu = e\tau/m$, parametrises how fast the electrons travel in an electric field since $v = \mu\mathcal{E}$. Typically in Si: the electron mobility is 0.15 ms^{-1}/Vm^{-1} and the hole mobility is 0.05 ms^{-1}/Vm^{-1} (because holes are heavier), so that in a field of 1 MVm^{-1} (1V across a 1 μm gap), electron travel about 10^5 ms^{-1}. This is one reason why transistors are shrinking as fast as technology allows them - there is an increase in the speed of electronics when keeping the same voltages because the electric fields increase, and hence the electrons' speed faster through the devices (in the absence of other effects). One of the problems though is that the mobility tends to be reduced when you accelerate the electrons too fast as they can gain enough energy to excite other electrons into higher energy states (by colliding with them).

The other component to carrier transport that must be considered occurs even when there is no applied electric field. This is diffusion of carriers from a region of high density to low density. You know already that a concentration gradient gives an electron flux $F_e = -D_e\frac{dn}{dx}$, producing a current density $j_e = eD_e\frac{dn}{dx}$, where D_e is the diffusion constant for electron carriers.

If we put a piece of n-doped semiconductor between the plates of a parallel plate capacitor then the charge $n(x)$ will redistribute so that in equilibrium there will be no net current flowing:
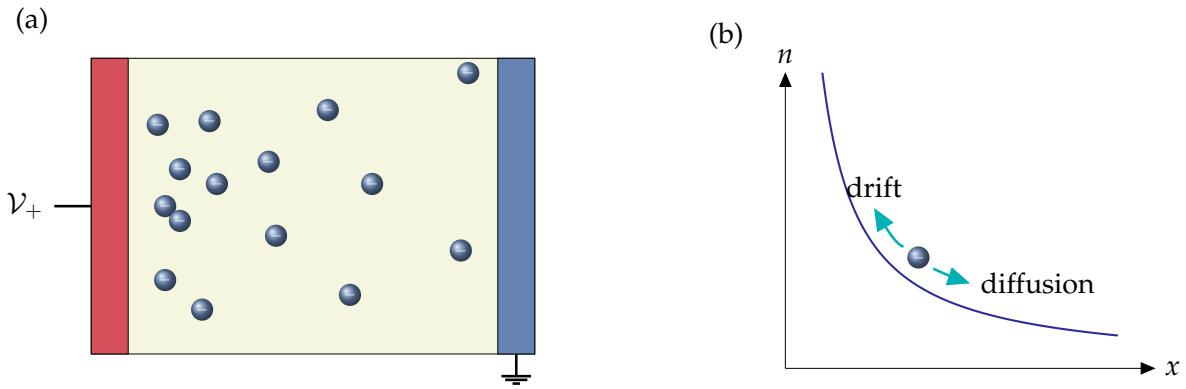
(a)

(b)

Figure 9.2: (a), Electrons in a crystal between parallel plates. (b) Change in electron density n with displacement.

If no net current is flowing in equilibrium then:

$$j_e = ne\mu_e \mathcal{E}(x) + eD_e \frac{dn}{dx} = 0 \tag{9.8}$$

Since $\mathcal{E}(x) = -d\mathcal{V}/dx$, we get:

$$ne\mu_e \frac{d\mathcal{V}}{dx} = eD_e \frac{dn}{dx} \tag{9.9}$$

Integrating:

$$n(x) = n_0 \exp\left(\frac{\mu_e}{D_e}\mathcal{V}(x)\right) \tag{9.10}$$

We should compare this to thermal equilibrium which says that

$$n(x) = n_0 \exp\left(\frac{e\mathcal{V}}{k_B T}\right) \tag{9.11}$$

so a surprising result drops out, called the Einstein relation:

$$\frac{\mu}{D} = \frac{e}{k_B T} \tag{9.12}$$

This is a very general result for any quasiparticle distribution obeying Maxwell-Boltzmann statistics (i.e. in thermal equilibrium). Now we have calculated how electrons move in electric fields we will see how to use them.

9.3 MOSFET

9.3.1 Gating Charge Distributions

One of the first benefits we get from using semiconductors is that we can change the carrier density at the Fermi energy by electrostatics.
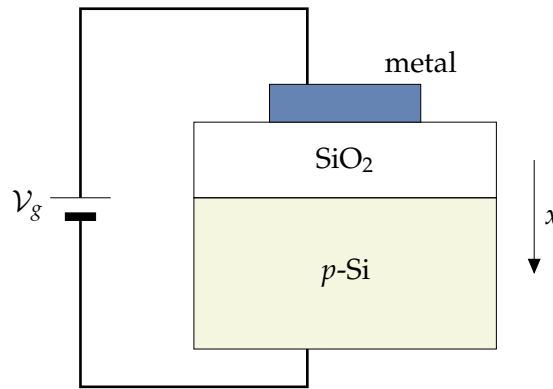
Figure 9.3: Basic metal-oxide-semiconductor structure that allows charge gating.

A thin layer of doped silicon is oxidised to form an insulator. With a metallic gate electrode deposited on top, we have the basis of a huge fraction of the trillions of transistors in the world! Let us take the doping as p-type, so that there are holes in the valence band. If a positive voltage is applied to the gate, then the electrostatic energy of holes under the gate increases. Assume that the interfaces are ideal and no charge is trapped in the insulator (this assumption is usually incorrect!). First let us calculate the hole concentration. Looking at the current perpendicular to the surface and in equilibrium:

$$j_h = pe\mu_h \mathcal{E}(x) - eD_h \frac{dp}{dx} = 0 \tag{9.13}$$

so as before:

$$-\frac{e}{k_B T} \frac{d\mathcal{V}}{dx} = \frac{1}{p} \frac{dp}{dx} \tag{9.14}$$

Integrating from the oxide/semiconductor interface into the bulk Si

$$-\frac{e\mathcal{V}_g}{k_B T} = \ln\left(\frac{p(0)}{p_0}\right) \tag{9.15}$$

$p(0)$ = hole concentration at interface.

p_0 = hole concentration in bulk semiconductor ($x = \infty$)

$\mathcal{V}_g$ = voltage across semiconductor

hence:

$$p(0) = p_0 \exp\left(-\frac{e\mathcal{V}_g}{k_B T}\right) \tag{9.16}$$

Holes are repelled from the surface by applying a positive gate voltage, as expected. Similarly for electrons, $n(0) = n_0 \exp\left(\frac{e\mathcal{V}_g}{k_B T}\right)$, thus the concentration of electrons is increased (they move towards the interface). We can check that the relation $np = n_0 p_0 = n_i^2$ is still preserved, since everything is in equilibrium.
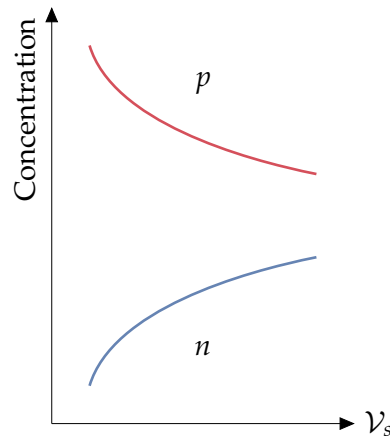
Figure 9.4: Concentration of electrons and holes with increasing bias.

Now consider a concentration of acceptors $N_a = 10^{21}$ m^{-3} (reasonable), so that $p_0 = 10^{21}$ m^{-3}, then $n_0 = n_i^2/p_0 \sim 2 \times 10^{11}$ m^{-3} is very small. But by putting a small voltage across the semiconductor the electron concentration can be increased!

$V_g = 0.1$V produces:

$p(0) = p_0 e^{-4} = 2 \times 10^{19}$ m^{-3}

$n(0) = n_0 e^{-4} = 10^{13}$ m^{-3}

Remember: $n_i = 1.5 \times 10^{16}$ m^{-3} for Si.

The first thing to note is that the holes have been pushed away from the surface but the electrons attracted do not make up for the decrease in positive charge. What dominates are the electrons which were trapped by the acceptors to liberate the holes, and this now forms a background region of negative charge. Thus the total charge density at the surface $= e(p - n - N_a) \sim -eN_a$.

The holes are only pushed away from the surface region if they are within the *depletion region*, x_d. We are now going to derive an estimate for x_d. We start off by approximating the charge distribution by $\rho = -eN_a$, $(x < x_d)$.

From this charge distribution it is easy to work out the potential in the material. You use Gauss' law:

$$\oint \mathbf{E} \cdot d\mathbf{S} = \frac{Q}{\varepsilon_r \varepsilon_0} \tag{9.17}$$

which gives (as in a capacitor)

$$\mathcal{E}A = eN_a(x - x_d)A/\varepsilon_r \varepsilon_0 \tag{9.18}$$

for $x < x_d$, since $\mathcal{E} = 0$ at $x = x_d$.
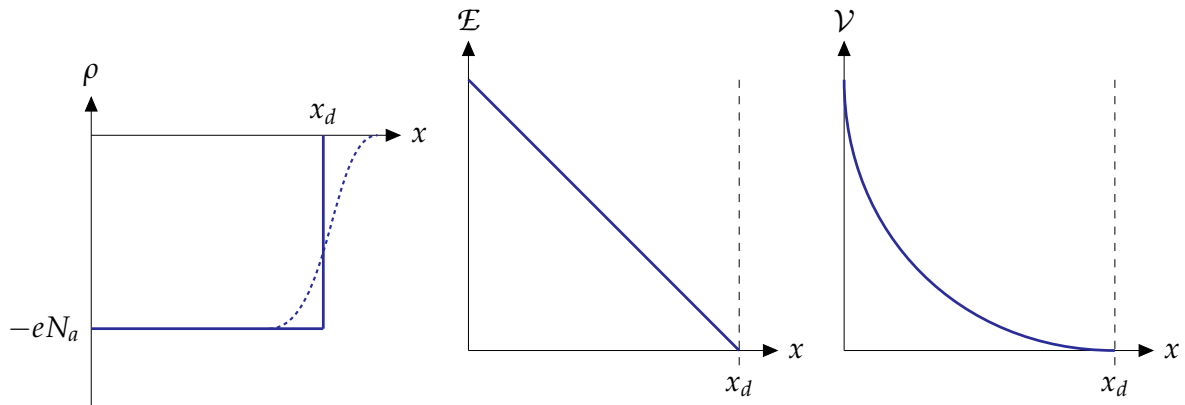
To get the potential we integrate:

Figure 9.5: Charge, field and potential distribution near a surface gate.

$$\mathcal{V}(x) = -\int_x^{x_d} \mathcal{E}(x)\, dx$$

$$= \frac{eN_a}{2\varepsilon_r\varepsilon_0}(x - x_d)^2 \tag{9.19}$$

At $x = x_d$, $\mathcal{V}(x) = 0$ as expected. The field from the contact has been *screened* by the mobile charge in the semiconductor. We also know that $\mathcal{V}(0) = \mathcal{V}_g$, so that (using typical values):

$$x_d = \sqrt{\frac{2\varepsilon_r\varepsilon_0\mathcal{V}_g}{eN_a}} \sim \sqrt{\frac{2.10.10^{-11}.10^{-1}}{10^{-19}.10^{21}}} \sim \mu\mathrm{m} \tag{9.20}$$

This equation shows how deep the *depletion width* is. In other words, this is how deep into the semiconductor that the voltage on the metal contact on top can influence. For higher doping concentrations, this depth becomes smaller. Our picture is changed slightly if charge can be stored at the interface between the insulating barrier and the Si since the potential at the top of the silicon is now less than applied gate potential (see Gauss's law again), so the switching effect is reduced.

9.3.2 Gating Devices

If we increase the gate voltage $\mathcal{V}_g$ enough, then the electron concentration at the surface of the Si will exceed the hole concentration. This 'gate' voltage is large enough so that the bottom of the valence band moves down enough to block holes moving underneath. Instead a layer of electrons collects right at the surface which can carry current between two 'metallic' contacts outside the gate region. Changing the gate voltage changes the electron density (just like restricting the water flow in a squashable plastic pipe), which can turn the transistor on and off.

This is impossible with a metallic conductor since although electrons would be repelled from the gate, there would still be a Fermi energy with empty electrons states above it that could carry a current. The semiconductor bandgap acts like a valve that allows us to turn off electron currents.
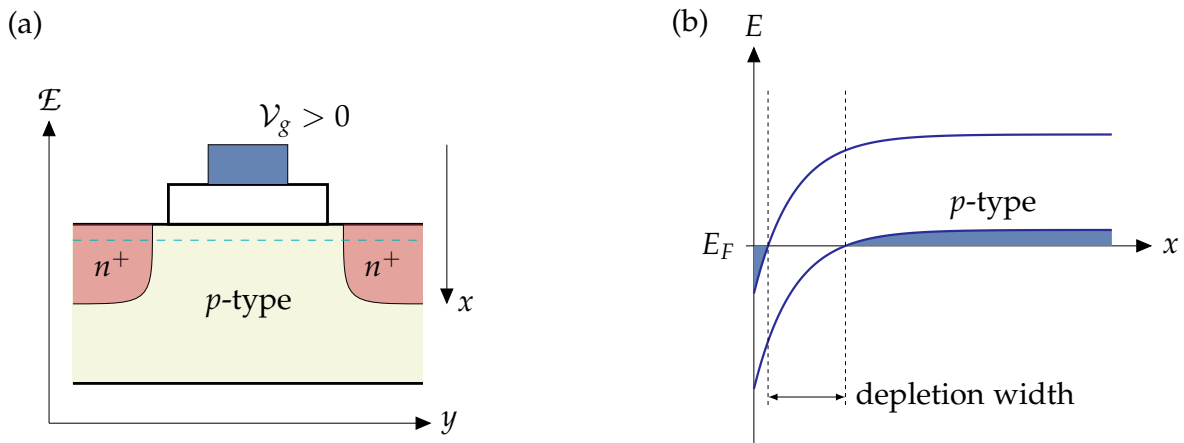
(a)

(b)

Figure 9.6: (a), Cross section of transistor. (b), Band bending of conduction and valence band edges due to the applied gate voltage.

This incredibly simple circuit is the basis of the computer's Dynamic Random Access Memory (DRAM), and most of the other logic circuits. Information is stored as a charge on the gate electrode which is written by a voltage pulse. Providing this charge does not leak away, it is possible to sense whether it is present or not by measuring the current in the semiconducting silicon underneath. What is very important about this device is that it can be made smaller and smaller and it still works well. Current transistors store about a million electrons, on an area about 200nm $\times$ 200nm.
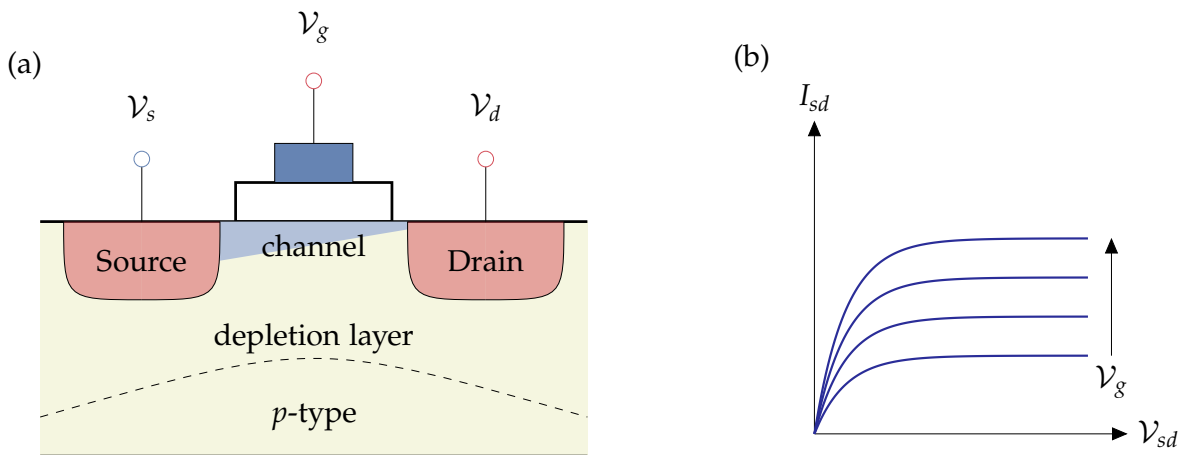
(a)

(b)

Figure 9.7: (a), Varying channel width along MOSFET due to varying relative gate voltage to channel. (b), Current-voltage characteristic of a MOSFET for different gate voltages.

We can bias ($\mathcal{V}_d$) one of the contacts positive so that electrons will be injected from the other contact (the source) and travel under the gate to be extracted at the drain electrode. This complicates matters since the extra positive potential on the drain reduces the voltage between the gate and Si at this end, reducing the attraction of electrons so that the depletion region takes a roughly triangular form.

The current through the channel depends on the electron mobility (at the surface) and density as well as the dimensions of the device (both its width and length).

As we look at the properties of semiconductors, you should realise that instead of accepting

the materials found as elements or compounds, the functionality emerges from growing and designing ways to manipulate the materials so it does what we want. It becomes a mixture of physics and technology since new basic physical principles can emerge from putting together the materials we know in different ways.

9.3.3   Size Scaling and Quantum Devices

As we discussed before, it is actually advantageous to reduce the size of all these semiconductor gates:

- larger fields mean larger drift velocity and faster operation

- shorter distances mean faster operation

- small areas mean lower capacitances mean faster operation

- Smaller size allows smaller fields, lower voltages and less power dissipation.

- Smaller size allows more gates / unit area: greater density, functionality

Disadvantages:

- smaller size: cannot reduce charge stored without risking errors.

- smaller size: new quantum effects start to become important

One of the capabilities of these new technologies is to produce potentials which can confine electrons to the same scale as their Fermi wavelength. These quantum effects can be seen in the MOSFET above.

9.4   Quantum Effects

Because of the bending of the bands by the gate voltage, the electrons are confined in a thin sheet, and we have to consider the quantum mechanics in the x direction. The electrons can all fall into the lowest of these quantum-confined energy levels but they are still able to move in the plane of the interface, so they form a 2-dimensional sheet.

Research into this 2D electron gas in similar structures is an extremely important and active area of physics as well as technology. This is mainly because the mobility of the electrons can be incredibly high - they are confined in sheets and thus there is less for them to scatter off as they drift in a field.

In some cases you can start to think of them like billiard balls in small devices with elastic walls. In other cases, by applying a magnetic field normal to the plane (x direction), the electrons orbit in circles (called *cyclotron* orbits) which further restricts their energies. You can see why this might be, from the parallel with the Bohr model of the atom. When an electron has travelled around one orbit it needs to be in phase with the original wavefunction. The frequency of the orbits is given by balancing the centripetal force with the force on the moving electron:
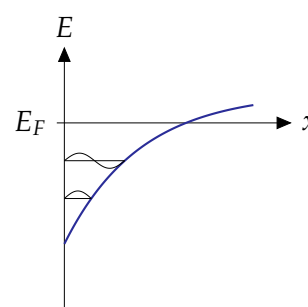

Figure 9.8:   Band bending due to quantum effects. 2D electron gas has electron energy levels with atomic (hydrogen-like) character.
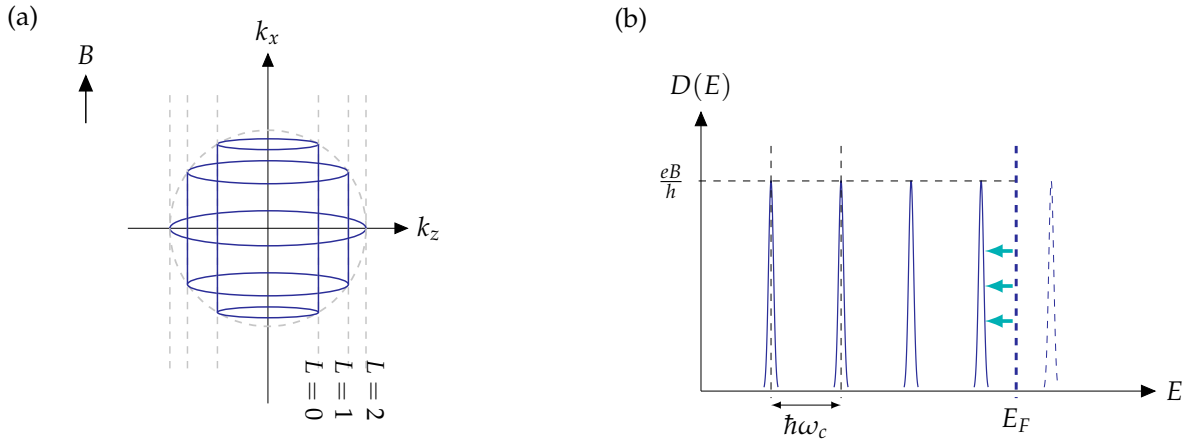
Figure 9.9: (a), Landau levels inscribed within the free electron Fermi sphere. (b), Density of states for quantum Hall effects. Allowed states in the quasi-continuum group into discrete Landau levels.

$$evB = \frac{m^* v^2}{r} \implies \frac{v}{r} = \omega_c = \frac{eB}{m^*} \tag{9.21}$$

Energy levels in the (k_y^2, k_z^2) plane are then quantised in accordance with a harmonic oscillator with frequency ω_c and results in a series of discrete quantised Landau levels with energy given by:

$$E_{yz}(L) = (L + \frac{1}{2})\hbar\omega_c, \tag{9.22}$$

where each energy state with quantum number $L \in \{0, 1, 2, 3, ...\}$ has the same degeneracy p. The energy E_{yz} is provided by the kinetic energy of the electrons in their plane, so that:

$$E(k_x, k_y) = \frac{\hbar^2}{2m^*}(k_y^2 + k_z^2) = E_{yz}(L) \tag{9.23}$$

This reduces the allowed values of $k_{y,z}$ and changes the density of states dramatically. The presence of a strong magnetic field gives rise to a system of concentric cylindrical Landau levels whose occupancy is enveloped by the otherwise spherical Fermi surface of a free electron gas (Fig. 9.9). The original quasi-continuous distribution of states have now condensed onto a series of discrete Landau levels, from both higher and lower energies. The area between each Landau ring is given by:

$$A_L = \pi \, d(k^2) = \frac{\pi 2m^* dE}{\hbar^2} = \frac{2\pi m^* \omega_c}{\hbar} \tag{9.24}$$

where $dE = \omega_c \hbar$ as each Landau level is separated in energy by this amount. The degeneracy of each Landau level is then given by:

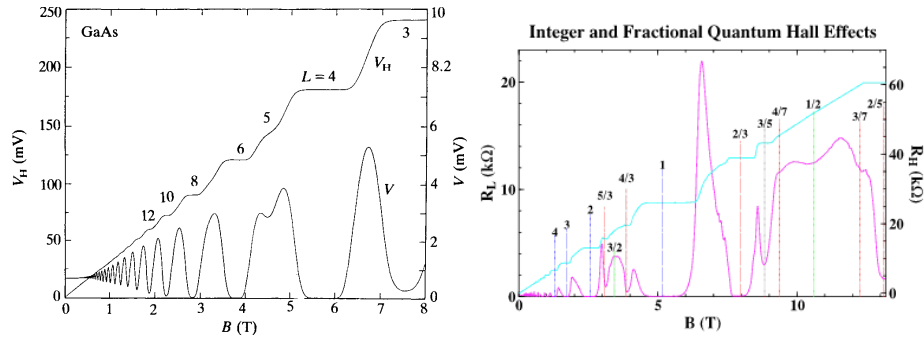$$p_L = A_L D(k) = \frac{eB}{\pi\hbar} \tag{9.25}$$

Figure 9.10: Integer and Fractional Quantum Hall Effects.

For all completely filled Landau levels (where $L_{\max}$ is the upper most occupied level), the total number of confined electrons is given by:

$$n_e = \sum p_L = p_L(L_{\max} + 1) = \frac{2eB}{h}(L_{\max} + 1) \tag{9.26}$$

Using the result for the transverse conductivity σ_{yz} of a free electron gas in a magnetic field[1] and considering equations 9.6,and 9.21 we can write:

$$\sigma_{yz} = -\frac{\sigma_e \omega_c \tau}{1 + \omega_c^2 \tau^2} \approx -\frac{\sigma_e}{\omega_c \tau} = -\frac{n_e e^2}{\omega_c m} = -\frac{n_e e}{B} \tag{9.27}$$

The resistivity is then given by:

$$\rho_{yz} = \frac{h}{2e^2} \frac{1}{L_{\max} + 1} \tag{9.28}$$

Equation 9.28 tells us that if the Fermi level lies exactly between two Landau levels, the transverse resistance is quantised in units of $h/2e^2$. In two dimensions, this resistance is identical to the Hall resistance. This is the *quantum Hall effect*, which shows different properties than we found for the normal Hall effect conduction in a thicker sample. As we have seen before, when the density of states has energy gaps, the electron transport is strongly affected. In this case, as the magnetic field B increases, the degeneracy p increases resulting an a greater number of electrons that fit into each quantised level. Each landau cylinder also expands with increasing B which effectively causes the Fermi energy to drop (as there are the same total number of electrons) and the number of occupied Landau levels decreases. When the Fermi energy is between the levels, then no current can be carried through the main sheet of the sample. But when the Fermi energy is centred at a level, electrons can flow and scatter into new states (in a very similar way to the 1D wires we discussed in Section 5.5.2). This behaviour radically changes the Hall voltage seen. The quantum levels are so clear (at low temperatures) that they have become the basis of a new current standard. Previously the force on coils balanced by masses were used, which extremely difficult experiments to do accurately. Resistances however can be accurately measured, leading from other measurements of h and e to a definition of current.

---

[1]See Myers box 6.1

Appendix A

Bloch's Theorem

Bloch's theorem states that "The wavefunctions $\psi(\mathbf{r})$ of electrons in crystals always have the property that if we observe the wavefunction about a lattice point, $\mathbf{R}$, it is the same as about the origin except for a phase shift which is given by multiplying the wavefunction by $e^{i\mathbf{k}\cdot\mathbf{R}}$."

That is to say that:

$$\psi(\mathbf{r} + \mathbf{R}) = \psi(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{R}} \tag{A.1}$$

Another way of writing this is:

$$\psi(\mathbf{r}) = u(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}} \tag{A.2}$$

where $u(\mathbf{r} + \mathbf{R}) = u(\mathbf{r})$ for lattice vectors $\mathbf{R}$.

Our proof proceeds as follows: We can define a translation operator $T_{\mathbf{R}}$ that shifts the argument of any function $f(\mathbf{r})$ by $\mathbf{R}$:

$$T_{\mathbf{R}}f(\mathbf{r}) = f(\mathbf{r} + \mathbf{R}) \tag{A.3}$$

Hamiltonian of crystal is periodic in $\mathbf{R}$ such that:

$$T_{\mathbf{R}}H\psi = H(\mathbf{r} + \mathbf{R})\ \psi(\mathbf{r} + \mathbf{R}) = H(\mathbf{r})\ \psi(\mathbf{r} + \mathbf{R}) = HT_{\mathbf{R}}\psi \tag{A.4}$$

or:

$$T_{\mathbf{R}}H = HT_{\mathbf{R}} \tag{A.5}$$

If operators commute then there exists a set of eigenfunctions for both. So we write eigenstate equations as:

$$H\psi = E\psi T_{\mathbf{R}}\psi = c(\mathbf{R})\psi \tag{A.6}$$

If we apply the translation operator twice, we have:

$$T_{\mathbf{R}'} T_{\mathbf{R}} \psi = c(\mathbf{R}) T_{\mathbf{R}'} \psi = c(\mathbf{R}) c(\mathbf{R}') \psi \tag{A.7}$$

But translations should also sum:

$$T_{\mathbf{R}'} T_{\mathbf{R}} \psi = T_{\mathbf{R}+\mathbf{R}'} \psi = c(\mathbf{R} + \mathbf{R}') \psi \tag{A.8}$$

Therefore we have:

$$c(\mathbf{R} + \mathbf{R}') \psi = c(\mathbf{R}) c(\mathbf{R}') \psi \tag{A.9}$$

Let $\mathbf{a}_i$ be three primitive lattice vectors. We can write:

$$c(\mathbf{a}_i) = e^{2\pi i x_i} \tag{A.10}$$

for a suitable x_i. If $\mathbf{R}$ is lattice vector given by $\mathbf{R} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3$, then:

$$c(\mathbf{R}) = c(n_1 \mathbf{a}_1) c(n_2 \mathbf{a}_2) c(n_3 \mathbf{a}_3) \tag{A.11}$$

But we can write this as:

$$c(\mathbf{R}) = e^{i \mathbf{k} \cdot \mathbf{R}} \tag{A.12}$$

where $\mathbf{k} = x_1 \mathbf{b}_1 + x_2 \mathbf{b}_2 + x_3 \mathbf{b}_3$ and $\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi \delta_{ij}$.

We therefore conclude that:

$$T_{\mathbf{R}} \psi = \psi(\mathbf{r} + \mathbf{R}) = c(\mathbf{R}) \psi = e^{i \mathbf{k} \cdot \mathbf{R}} \psi(\mathbf{r}) \tag{A.13}$$

□

Appendix B

The Brillouin Function

At thermal equilibrium the number of dipoles with various orientations and energies is given by the Maxwell-Boltzmann distribution. The partition function can be written as:

$$Z = \sum_{m_J=-J}^{J} e^{-\eta m_J}, \qquad \eta = \frac{g \mu_B B}{k_B T} \tag{B.1}$$

The Helmholtz free energy is given by:

$$F = -k_B T \ln(Z) \tag{B.2}$$

The unit magnetisation is defined as:

$$M \equiv - \left. \frac{\partial F}{\partial B} \right|_{S,V} \tag{B.3}$$

Substituting the Helmholtz equation we arrive at:

$$M = \frac{k_B T}{Z} \frac{\partial Z}{\partial B} = \frac{\sum_{m_J=-J}^{J} m_J g \mu_B \ e^{-\eta m_J}}{\sum_{m_J=-J}^{J} \ e^{-\eta m_J}} \tag{B.4}$$

We write the partition function as a sum of a geometric series:

$$
\begin{aligned}
Z &= \frac{e^{\eta J} - e^{-\eta(J+1)}}{1 - e^{-\eta}} \\
&= \frac{e^{\eta(J+1/2)} - e^{-\eta(J+1/2)}}{e^{\eta/2} - e^{-\eta/2}} \tag{B.5} \\
&= \frac{\sinh\left\{\eta\left(J + \frac{1}{2}\right)\right\}}{\sinh\left\{\frac{\eta}{2}\right\}} \tag{B.6}
\end{aligned}
$$

Also we know that:

$$\begin{aligned}
\frac{\partial Z}{\partial B} &= \frac{\partial Z}{\partial \eta}\frac{d\eta}{dB} \\
&= \frac{g\mu_B}{k_B T}\left\{ \frac{(J+1/2)\cosh(\eta(J+1/2))}{\sinh(\eta/2)} \right. \\
&\quad \left. - \frac{\frac{1}{2}\coth(\eta/2)\sinh(\eta(J+1/2))}{\sinh(\eta/2)} \right\}
\end{aligned}$$ (B.7)

Substituting yields:

$$M = \frac{k_B T}{Z}\frac{\partial Z}{\partial B} = g\mu B\left\{ (J+1/2)\coth(\eta(J+1/2)) - \frac{1}{2}\coth(\eta/2) \right\}$$ (B.8)

If we now make the substitution for η such that $\eta = \alpha/J$ we obtain:

$$\begin{aligned}
M &= g\mu_B J\left\{ B_J(\alpha) \right\} \\
B_J(\alpha) &= \frac{2J+1}{2J}\coth\left(\frac{\alpha(2J+1)}{2J} \right) - \frac{1}{2J}\coth\left(\frac{\alpha}{2J} \right)
\end{aligned}$$ (B.9)

2019 Newton, M.C.

Appendix C

The Exchange Interaction

The exchange energy is the manifestation of the permutation symmetry that is required of the many-particle wave function of identical particles. Exchange effects can occur in all elementary particles and can take a differing form in each case. For the case of electrons, their indistinguishability demands a wave function that is antisymmetric with the exchange of the coordinates of two electrons (including both spatial and spin coordinates). We can illustrate the origin of exchange effects by solving the two-electron problem (e.g.: Helium atom), while ignoring the electron-electron Coulomb interaction. In doing so, we will demonstrate that although the Coulomb interaction between the electrons and nuclei does not depend explicitly on the electron spin, the energy of the electrons depends on their relative spin. The spin independent two-electron Hamiltonian can be written as:

$$H = -\frac{\hbar^2}{2m}\left\{\nabla_1^2 + \nabla_2^2\right\} + U \tag{C.1}$$

Where the subscripts 1 and 2 are with respect to electrons at positions $\mathbf{r}_1$ and $\mathbf{r}_2$ and U is the potential energy of the system. If the single particle coordinate wave functions are $\psi_a(\mathbf{r})$ and $\psi_b(\mathbf{r})$, the two possible eigenvectors of the two-electron system take either symmetric (S) or anti-symmetric (A) coordinates:

$$\Psi_S(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}}\left(\psi_a(\mathbf{r}_1)\psi_b(\mathbf{r}_2) + \psi_a(\mathbf{r}_2)\psi_b(\mathbf{r}_1)\right) \tag{C.2a}$$

$$\Psi_A(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}}\left(\psi_a(\mathbf{r}_1)\psi_b(\mathbf{r}_2) - \psi_a(\mathbf{r}_2)\psi_b(\mathbf{r}_1)\right) \tag{C.2b}$$

The expectation value of the interaction energy in each case is then given by:

$$\langle U_{\mathrm{S}} \rangle = \int \Psi_{\mathrm{S}}^{*} \, U \, \Psi_{\mathrm{S}} \mathrm{d}\mathbf{r} = C_{U} + J_{U} \tag{C.3a}$$

$$\langle U_{\mathrm{A}} \rangle = \int \Psi_{\mathrm{A}}^{*} \, U \, \Psi_{\mathrm{A}} \mathrm{d}\mathbf{r} = C_{U} - J_{U} \tag{C.3b}$$

$$C_{U} = \iint \psi_{a}(\mathbf{r}_1) \psi_{b}(\mathbf{r}_2) \, U \, \psi_{a}^{*}(\mathbf{r}_1) \psi_{b}^{*}(\mathbf{r}_2) \, \mathrm{d}\mathbf{r}_1 \, \mathrm{d}\mathbf{r}_2 \tag{C.4a}$$

$$J_{U} = \iint \psi_{a}(\mathbf{r}_1) \psi_{b}^{*}(\mathbf{r}_1) \, U \, \psi_{a}^{*}(\mathbf{r}_2) \psi_{b}(\mathbf{r}_2) \, \mathrm{d}\mathbf{r}_1 \, \mathrm{d}\mathbf{r}_2 \tag{C.4b}$$

If we assume that the single electron wave functions are non-negative and the potential U is also non-negative (Coulomb interaction), then both C_{U} and J_{U} take on positive values. Recall that in order to construct an antisymmetric wave function, we can use either a symmetric spatial coordinate wave function with an antisymmetric spin wave function or an anti-symmetric spatial coordinate wave function with a symmetric spin wave function. It then follows that the energy of the symmetric (singlet) state (anti-parallel spins) $\langle U_{\mathrm{S}} \rangle$ is higher than that of the antisymmetric state (parallel spins) $\langle U_{\mathrm{A}} \rangle$. In an isolated atom, the exchange energy arises only between electrons with the same spin and results in an overall reduction in the total energy of the atom. This is the reason why the electrons are typically arranged to produce a total spin that is the largest possible (Hund's rules).

We can consider the force between electrons in a general way as $F = -\partial U / \partial R$, where R is the spatial separation. As the gradient of C_{U} and J_{U} are both negative, the force due to the exchange interaction J_{U} is positive (repulsive) for the case of symmetric coordinates (anti-parallel spins) and negative (attractive) for anti-symmetric coordinates (parallel spins).