Quantum Condensed Matter Physics Part II Lent/Easter 2006

P B Littlewood

January 6, 2006

Contents

1	Intr	roduction					
	1.1	Theories and models $\ldots \ldots \ldots$	9				
	1.2	"The Properties of Matter"	10				
	1.3	Collective phenomena and emergent properties	11				
2	The	e variety of condensed matter	13				
	2.1	Types of solids	13				
	2.2	The binding of crystals	13				
	2.3	Complex matter	20				
3	Fer	mi and Bose gases	25				
	3.1	Free electron gas in three-dimensions	25				
	3.2	Fermi surface, and density of states	26				
	3.3	Thermal properties of the electron gas $\ldots \ldots \ldots \ldots \ldots$	27				
	3.4	Lattice dynamics and phonons $\hfill \ldots \hfill \hfill \ldots \hfill \ldots \hfill \ldots \hfill \ldots \hfill \ldots \hfill \ldots \$	29				
	3.5	Lattice specific heat	36				
4	\mathbf{Per}	iodic Solids and Diffraction	39				
	4.1	The description of periodic solids $\ldots \ldots \ldots \ldots \ldots \ldots \ldots$	39				
	4.2	The reciprocal lattice and diffraction $\hdots \hdots \hd$	41				
	4.3	Diffraction conditions and Brillouin zones	43				
5	Eleo	ctronic structure 1	45				
	5.1	Tight binding: Linear combination of atomic orbitals	45				
	5.2	Diatomic molecule revisited $\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$	45				
	5.3	Linear chain	46				

CONTENTS

	5.4	Bloch's theorem	47				
	5.5	Linear chain revisited $\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$	49				
	5.6	LCAO method in general $\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$	49				
	5.7	Periodic boundary conditions and counting states in 3 dimensions	51				
6	Eleo	ctronic structure 2	53				
	6.1	Nearly free electrons	53				
	6.2	The Schrödinger equation in momentum space \ldots	53				
	6.3	One-dimensional chain	54				
	6.4	Pseudopotential	58				
7	Ban	d theory of metals and insulators	61				
	7.1	Bands and Brillouin zones	61				
	7.2	Metals and insulators in band theory	62				
	7.3	Examples of band structures	66				
	7.4	Probing the band structure and density of states	70				
8	Elec	ctrodynamics of metals	81				
	8.1	Screening and Thomas-Fermi theory	81				
	8.2	Plasma oscillations	85				
	8.3	Optical conductivity of metals	86				
9	Sen	niclassical dynamics	91				
	9.1	Wavepackets and equations of motion $\ldots \ldots \ldots \ldots \ldots$	91				
	9.2	Electrons and holes in semiconductors \hdots	95				
	9.3	Scattering and electrical conductivity of metals $\ldots \ldots \ldots$	97				
	9.4	Transport in electrical and magnetic fields	99				
10	Sen	niconductors	103				
	10.1	Semiconductor band structure	103				
	10.2	Intrinsic carrier concentration	106				
	10.3	Extrinsic semiconductors	108				
11 Semiconductor devices 11							
	11.1	Metal - semiconductor contact	111				
	11.2	p-n junction	112				

4

CONTENTS

11.3	Field effect transistor	•		•	•	 	 •	•	•	119
11.4	Compound semiconductor heterostructures	3		•	•	 				120

Preface

Books

There are many good books on solid state and condensed matter physics, but the subject is rich and diverse enough that each of these contains both much more and much less than the topics covered in this course. The two classic textbooks are Kittel, and Ashcroft and Mermin. These are both at the correct level of the course, and have the virtue of clear exposition, many examples, and lots of experimental data. Slightly more concise, though in places a little more formal is Ziman. Grosso and Parravicini has a somewhat wider coverage of material, but much of it goes well beyond the level of detail required for this course. Marder is at about the right level (though again with more detail than we shall need), and has a nice blend of quantum properties with statistical and classical properties. OUP have recently issued a series of short texts in condensed matter physics. They are more detailed than needed for this course, but are quite accessible and excellent for reference. The most relevant for this course is Singleton.

- C.Kittel, Introduction to Solid State Physics, 7th edition, Wiley, NY, 1996.
- N.W.Ashcroft and N.D.Mermin, *Solid State Physics*, Holt-Saunders International Editions, 1976.
- J.M.Ziman, Principles of the Theory of Solids, CUP, Cambridge, 1972.
- J. Singleton, Band Theory and the Electronic Properties of Solds, OUP 2001.
- M.P. Marder, *Condensed Matter Physics*, Wiley, NY, 2000. Covers both quantum matter and mechanical properties.
- G.Grosso and G.P.Parravicini, *Solid State Physics*, AP, NY, 2000. A wide coverage of material, very bandstructure oriented, very detailed.
- A very good book, though with a focus on statistical and "soft" condensed matter that makes it not so relevant for this course, is P.M.Chaikin and T.Lubensky, *Principles of Condensed Matter Physics*, CUP, Cambridge, 1995.

These notes

These notes are designed as a complement to the lectures, and as a complement to reference books. They are not to be memorised for examination: often they include detailed derivations that are there to satisfy the curious, for completeness, and for background. The lectures will be presented using more qualitative and physical descriptions.

CONTENTS

In a few places, and particularly where I shall be discussing material that is not easy to find collected in textbooks, the notes are much more lengthy. Material which is *explicitly* non-examinable is placed in small type; but in general, no detailed derivations will be required for examination. You may find it worthwhile, however, to work through some of this at least once.

Problems are placed both within the text and at the ends of the chapters. They vary from the straightforward to the complex, and especially mathematical ones are given a warning asterisk. Problems of such type will not appear in the examination.

CONTENTS

Chapter 1

Introduction

1.1 Theories and models in condensed matter physics

Solid state physics is concerned with the abundance of properties that arise when atoms are amalgamated together. Much of what we think of as "core physics" is deliberately reductionist; we look for the very simplest unified description of a basic phenomenon, and the progress of much of basic physics has always been a progress toward grander unified theories, each of which is simpler (at least in concept) than the previous generation.

Condensed matter physics is not like this. The Hamiltonian is not in doubt - it is the Schrödinger equation for the many particle system:

$$H_{elec} = -\sum_{i} \frac{\hbar^{2}}{2m} \nabla_{i}^{2} + \sum_{I} \frac{P_{I}^{2}}{2M_{I}} + \sum_{i,I} \frac{Z_{I}e^{2}}{|\mathbf{r}_{i} - \mathbf{R}_{I}|} + \frac{1}{2} \sum_{i \neq j} \frac{e^{2}}{|\mathbf{r}_{i} - \mathbf{r}_{j}|} + \frac{1}{2} \sum_{I \neq J} \frac{Z_{I}Z_{J}e^{2}}{|\mathbf{R}_{i} - \mathbf{R}_{j}|} ,$$
(1.1)

where the \mathbf{r}_i , \mathbf{R}_I label the coordinates of the electrons and the ions respectively, Z_I , M_I are the nuclear charge and mass. The terms in Eq. (1.1) represent, in order, the kinetic energy of the electrons, the kinetic energy of the nuclei, and the Coulomb interaction between electron and nucleus, electron and electron, and between nucleus and nucleus. In some sense, a complete theory of solids would be to solve the Schrodinger equation and then apply all the standard methods of statistical physics to determine thermodynamic and physical properties. From this point of view, there is no "fundamental" theory to be done, although the calculations may indeed be complex (and in fact, impossible to perform exactly for solids with macroscopic numbers of atoms). Because an exact solution for a macroscopic number of atoms is impossible, we have to treat Eq. (1.1) by a sequence of approximations (for example, perhaps fixing the ions in place, or neglecting electron-electron interactions) that will make the problem tractable.

This view of condensed matter physics as a series of approximations that is widely held and severely incomplete. Suppose for a moment that we could solve the full Hamiltonian, and we would then have a wavefunction describing some 10^{23} particles that contained all of the physics of solids. Writing the solution down would be hard enough, but comprehending its meaning would be beyond us. Condensed matter physics is about phenomena, from the mundane (why is glass transparent), to the exotic (why does ³He become a superfluid). There are a host of physical phenomena to be understood, and their explanation must involve more than just detailed calculation.

Understanding a phenomenon involves building the simplest possible model that explains it, but the models are more than just approximations to Eq. (1.1). Models, and the theories which they give rise to, elucidate paradigms and develop concepts that are obscured by the complexity of the full Hamiltonian. The surprise about condensed matter physics is that there are so many *different* theories that can arise from such an unprepossessing Hamiltonian as Eq. (1.1).

1.2 "The Properties of Matter"

A venerable route to condensed matter physics, and one followed by almost all textbooks, is to find ways of making approximate calculations based on the full Schrödinger equation for the solid. Making approximate, but quantitative calculations of the physical properties of solids has been one of the enduring agendas of condensed matter physics and the methods have acquired increasing sophistication over the years. We would like to understand the cohesion of solids - why it is, for example that mercury is a liquid at room temperature, while tungsten is refractory. We wish to understand electrical and optical properties – why graphite is a soft semi-metal but diamond a hard insulator, and why GaAs is suitable for making a semiconductor laser, but Si is not. Why is it that some materials are ferromagnetic, and indeed why is it that transition metals are often magnetic but simple s-p bonded metals never? We would like to understand chemical trends in different classes of materials – how properties vary smoothly (or not) across the periodic table. These, and many other physical properties we now know how to calculate with considerable accuracy by sophisticated computational techniques, but more importantly (and especially for the purposes of this course) we can understand the behaviour straightforwardly, and describe the physical properties in a natural fashion.

To get this understanding we need to develop the basic machinery of the quantum mechanics of periodic structures, especially the concept of electronic bandstructure describing the dispersion relation between the electron's energy and momentum. We also need to understand how the largest effects of interactions between electrons can be subsumed into averaged effective interactions between independent quasiparticles and the background medium. A large part of this course will be to set up this fundamental machinery.

This is a tidy scheme, but it will get us only part way to the goal. It will generate for us a landscape upon which we can build new models and new theories.

1.3 Collective phenomena and emergent properties

There is another view of condensed matter physics which we shall also explore, that is less concerned with calculation and more concerned with phenomena per se. The distinguishing character of solid state systems is that they exhibit *collective* phenomena, that are properties of macroscopic systems and that exist only on account of the many-degree-of-freedom nature of the system.

A familiar example is a phase transition (between liquid and solid, say) which is a concept that can only apply to a macroscopic ensemble. We are so used to phase transitions that we rarely wonder why when water is cooled down it does not just get "thicker" and more viscous (and this actually happens to a glass).

Condensed matter systems have collective modes that are a consequence of their order; both a solid and a liquid support longitudinal sound waves, but a solid that has a nonzero shear stiffness has also transverse sound modes. In fact the existence of shear waves we might choose to define as the characteristic feature distinguishing a solid from a liquid or gas. We can say that solidity is a *broken symmetry* (with the symmetry being broken that of translational invariance); because of the broken symmetry, there is a new collective mode (the shear wave). Because of quantum mechanics, the waves are necessarily quantised as phonons, and they are a true quantum particle, with Bose statistics, that interact with each other (due to anharmonicity) and also with other excitations in the solid. This idea, that a broken symmetry can generate new particles, is one of the central notions of condensed matter physics – and of course of particle physics too.

A different example is the behaviour of electrons in a semiconductor. If one adds an electron into the conduction band of a semiconductor it behaves like a particle of charge -|e|, but a mass different from the free electron mass due to the interaction with the lattice of positively charge ions as well as all the other electrons in the solid. But we know that if we remove an electron from the valence band of the semiconductor, it acts as a *hole* of charge +|e|; the hole is in fact a collective excitation of the remaining 10^{23} or so electrons in the valence band, but it is a much more convenient and accurate description to think of it as a new fermionic *quasi*-particle as an excitation about the ground state of the solid. The electrons and holes, being oppositely charged, can even bind together to form an exciton - the analog of the hydrogen atom (or more directly positronium), which however has a binding energy considerably reduced from hydrogen, because the Coulomb interaction is screened by the dielectric constant of the solid, and the electron and hole masses are different from the electron and proton in free space.

The solid is a new "vacuum", inhabited by quantum particles with properties which may be renormalised from those in free space (e.g. photons, electrons) or may be entirely new, as in the case of phonons, plasmons (longitudinal charge oscillations), magnons (waves of spin excitation in a magnet), etc. In contrast to the physical vacuum, there are different classes of condensed matter systems which have different kinds of vacua, and different kinds of excitations. Many of these new excitations arise because of some "broken" symmetry , for example, magnetism implies the existence of spin waves, and solidity implies the existence of shear waves. Some of these phenomena – superconductivity, superfluidity, and the quantum Hall effect come to mind – are remarkable and hardly intuitive. They were discovered by experiment; it seems unlikely that they would ever have been uncovered by an exercise of pure cerebration starting with the Schrodinger equation for 10^{20} particles.

Solid state systems consist of a hierarchy of processes, moving from high energy to low; on the scale of electron volts per atom are determined the cohesive energy of the solid, (usually) the crystal structure, whether the material is transparent or not to visible light, whether the electrons are (locally) magnetically polarised, and so on. But after this basic landscape is determined, many further phenomena develop on energy scales measured in meV that correspond to thermal energies at room temperature and below. The energy scales that determine magnetism, superconductivity, etc. are usually several orders of magnitude smaller than cohesive energies, and the accuracy required of an *ab initio* calculation would be prohibitive to explain them. Although all condensed matter phenomena are undoubtedly to be found within the Schrödinger equation, they are not transparently derived from it, and it is of course better to start with specific models that incorporate the key physics; we shall see many of them. These models will usually be simply of interactions between excitations of the solid, with sets of parameters to describe them - parameters which are usually estimated, or derived from experiment.

Chapter 2

The variety of condensed matter

2.1 Types of solids

What holds a solid together? Cohesion is ultimately produced by the electrostatic interaction between the nuclei and the electrons, but depending on the particular atomic structure the types of solids can be very different.

2.2 The binding of crystals

Inert gases

The inert gases have filled electron shells and large ionisation energies. Consequently, the electronic configuration in the solid is close to that of separated atoms. Since the atoms are neutral, the interaction between them is weak, and the leading attractive force at large distances comes from the van der Waals interaction, which gives an attractive potential proportional to $1/R^6$.

This form can be loosely derived by thinking of an atom as an oscillator, with the electron cloud fluctuating around the nucleus as if on a spring.

The centre of the motion lies on top of the atom, but if the cloud is displaced, there will be a small dipole induced, say p_1 . Such displacements happen as a result of zero-point motion of the electron cloud in the potential of the nucleus.

A distance R away from the atom there is now an induced electric field $\propto p_1/R^3$. A second atom placed at this point will then have a dipole induced by the electric field of the first: $p_2 \propto \alpha p_1/R^3$, where α is the atomic polarizability.

The second dipole induces an electric field at the first, which is now

$$E_1 \propto p_2/R^3 \propto \alpha p_1/R^6. \tag{2.1}$$

The energy of the system is then changed by an amount

$$\Delta U = \langle -p_1 \cdot E_1 \rangle \propto -\alpha \langle p_1^2 \rangle / R^6.$$
(2.2)

Notice that it depends on the expectation value of the square of the dipole moment $\langle p_1^2 \rangle$, which is non-zero, and not the square of the expectation value $\langle p_1 \rangle^2$, which would be zero.



Figure 2.1: Two dipoles represent model atoms that are arranged along a line, with the positive charges (+e) fixed at the positions 0, R, and the negative charges (-e) at the points $x_1, R + x_2$.

Qu.2.1 Interacting harmonic oscillators That van der Waals forces are in fact a consequence of the change in zero-point energy of coupled oscillators can be seen by the following toy example.

We assume $R \gg x_1, x_2$. The dipoles are connected by springs so that the Hamiltonian of the independent atoms is

$$H_0 = \frac{1}{2m}\hat{p}_1^2 + \frac{m\omega_0^2}{2}x_1^2 + \frac{1}{2m}\hat{p}_2^2 + \frac{m\omega_0^2}{2}x_2^2$$
(2.3)

where \hat{p}_i are the momenta, and ω_0 the lowest eigenfrequency.

Show that the interaction Hamiltonian is approximately

$$H_1 \approx -\frac{e^2}{2\pi\epsilon_0 R^3} x_1 x_2 \tag{2.4}$$

By making the transformation to symmetric and antisymmetric normal modes, $x_s = (x_1 + x_2)/\sqrt{2}$, $x_a = (x_1 - x_2)/\sqrt{2}$ $p_s = (p_1 + p_2)/\sqrt{2}$, $p_a = (p_1 - p_2)/\sqrt{2}$ show that the Hamiltonian can be rewritten as

$$H = H_0 + H_1 \approx \frac{1}{2m}\hat{p}_s^2 + \frac{m\omega_s^2}{2}x_s^2 + \frac{1}{2m}\hat{p}_a^2 + \frac{m\omega_a^2}{2}x_a^2$$
(2.5)

and determine ω_s and ω_a .

14

2.2. THE BINDING OF CRYSTALS

Hence show that the zero point energy of the system is

$$U_0 = \hbar \omega_0 \left[1 - \frac{1}{2} \left(\frac{e^2}{4\pi \epsilon_o R^3 m \omega_0^2} \right)^2 + \dots \right]$$
(2.6)

If the atoms move together so that the electron charge distributions begin to overlap, repulsive forces come into play. While there is of course a contribution from the direct electrostatic repulsion of the electrons, more important is the Pauli exclusion principle that prevents two electrons having their quantum numbers equal. The effect of Pauli exclusion can be seen by an extreme example, of overlapping two Hydrogen atoms entirely, with the electrons for simplicity assumed to be in the same spin state. In this case, while two separated atoms may be both in the 1S ground state, the combined molecule must have a configuration 1s2s, and thus is higher by the promotion energy.

Calculations of the repulsive interaction are complex but the answer is clearly short-ranged. They are often modelled empirically by an exponential form e^{-R/R_o} , or a power law with a large power. A commonly used empirical form to fit experimental data on inert gases is the Lennard-Jones potential

$$U(R) = -\frac{A}{R^6} + \frac{B}{R^{12}}$$
(2.7)

with A and B atomic constants obtained from gas-phase data.

With the exception of He, the rare gases from close-packed (face-centered cubic) solids with a small cohesive energy, and low melting temperatures. Helium is special because zero-point motion of these light atoms is substantial enough that they do not solidify at zero pressure down to the absolute zero of temperature. The quantum fluids ${}^{3}He$ and ${}^{4}He$ have a number of extraordinary properties, including superfluidity.

Ionic Crystals

Given the stability of the electronic configurations of a rare gas, atoms that are close to a filled shell will have a tendency to lose or gain electrons to fill the shell.

- The energy for the reaction $M > M^+ + e^-$ in the gas phase is called the ionization energy I.
- The energy for the reaction $X + e^- > X^-$ in the gas phase is called the electron affinity A.
- The cohesion of an ionic molecule can overcome the energy cost I + A by the electrostatic attraction, e^2/R

• In a solid, the electrostatic interaction energy for a diatomic crystal¹ is

$$U_{electrostatic} = \frac{1}{2} \sum_{i} \sum_{j} U_{ij}$$
(2.8)

where $U_{ij} = \pm q^2/R_{ij}$ is the sum of all Coulomb forces between ions. If the system is on a regular lattice of lattice constant R, then we write the sum

$$U_{electrostatic} = -\frac{1}{2} \frac{\alpha_M q^2}{R} \tag{2.9}$$

where α_M is a dimensionless constant that depends only on the crystal structure.

- The evaluation of α_M is tricky, because the sum converges slowly. Three common crystal structures are NaCl ($\alpha_M = 1.7476$), CsCl (1.7627), and cubic ZnS or Zincblende (1.6381).
- To the attractive Madelung term must be added the repulsive short range force, and we now have the added caveat that ions have different sizes, explaining why NaCl has the rocksalt structure, despite the better electrostatic energy of the CsCl structure.

Covalent crystals

The covalent bond is the electron pair or single bond of chemistry.

Model Hydrogen. Two overlapping atomic orbitals on identical neighbouring atoms will hybridise. Because the Hamiltonian must be symmetric about a point centered between the ions then the eigenstates must have either even or odd parity about this center. If we have a simple system of two one electron atoms - model hydrogen - which can be approximated by a basis of atomic states $\phi(r-R)$ (assumed real) centered on the nucleus R, then two states of even and odd parity are

$$\psi_{\pm}(r) = \phi(r - R_1) \pm \phi(r - R_2) \tag{2.10}$$

 ψ_+ has a substantial probability density between the atoms, where ψ_- has a node. Consequently, for an attractive potential $E_+ < E_-$, and the lower (bonding) state will be filled with two electrons of opposite spin. The antibonding state ψ_{-} is separated by an energy gap $E_{g} = E_{-} - E_{+}$ and will be unfilled. The cohesive energy is then approximately equal to the gap E_q^{-2}

Covalent semiconductors. If we have only s-electrons, we clearly make molecules first, and then a weakly bound molecular solid, as in H_2 . Using p,

16

 $^{^1\}mathrm{Beware}$ the factor of 1/2, which avoids double counting the interaction energy. The energy of a single ion i due to interaction with all the other ions is $U_i = \sum_{i \neq i} U_{ij}$; the total energy is $\frac{1}{2}\sum_{i}U_{i}$ ²Actually twice (two electrons) half the gap, if we assume that $E_{\pm} = E_{atom} \pm \frac{1}{2}E_{g}$



Figure 2.2: Tetrahedral bonding in the diamond structure. The zincblende structure is the same but with two different atoms per unit cell

d, orbitals, we may however make *directed* bonds, with the classic case being the sp^3 hybrid orbitals of C, Si, and Ge. These are constructed by hybrid orbitals $s+p_x+p_y+p_z+3$ other equivalent combinations, to make new orbitals that point in the four tetrahedral directions: (111), ($\overline{1}\overline{1}1$), ($\overline{1}1\overline{1}$), ($1\overline{1}\overline{1}$). These directed orbitals make bonds with neighbours in these tetrahedral directions, with each atom donating one electron. The open tetrahedral network is the familiar diamond structure of C, Si and Ge.

Ionic semiconductors. In GaAs and cubic ZnS the total electron number from the pair of atoms satisfies the "octet" rule, and they have the identical tetrahedral arrangement of diamond, but with the atoms alternating. This is called the *zinclende* structure. The cohesion in these crystals is now part ionic and part covalent. There is another locally tetrahedral arrangement called *wurtzite* which has a hexagonal lattice, favoured for more ionic systems. With increasing ionic components to the bonding, the structures change to reflect the ionicity: group IV Ge (diamond), III-V GaAs (Zincblende), II-VI ZnS(zincblende or wurtzite), II-VI CdSe (wurtzite), and I-VII NaCl (rocksalt).

Qu.2.2 Diatomic molecule This is a simple problem to illustrate the physics of a diatomic molecule (See Fig. 2.3). It also provides an elementary example of the *Linear Combination of Atomic Orbitals* or LCAO method, which we shall be using later to describe extended solids.

We restrict the basis of states to just the ground state of each atom in isolation, whereas of course an accurate solution would require a complete set of states that of necessity would include all the excited states of the atoms. The basis set consists of two states |1> and |2> that satisfy

$$H_1|1> = E_1|1> (2.11)$$

$$H_2|2> = E_2|2>$$
 (2.12)

and we look for solutions

$$|\psi\rangle = c_1|1\rangle + c_2|2\rangle \tag{2.13}$$

Neglecting the direct matrix elements <1|2> for simplicity (these are easily included if necessary), derive the matrix equation for the wavefunctions and eigenvalues

$$\begin{pmatrix} H_{11} - E & H_{12} \\ H_{21} & H_{22} - E \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = 0$$
(2.14)

where the matrix elements are of two kinds: Onsite, or *crystal field* terms

$$H_{11} = \langle 1 | T + V_1 + V_2 | 1 \rangle = E_1 + \langle 1 | V_2 | 1 \rangle = E_1$$
(2.15)

Offsite, or hopping terms³

$$H_{12} = \langle 1 | T + V_1 + V_2 | 2 \rangle = t \tag{2.16}$$

Solve for the wavefunctions and eigenvalues, for t < 0.

Sketch the wavefunctions and charge densities for the lower and upper states, in the cases of (a) identical atoms $\tilde{E}_1 = \tilde{E}_2$, and (b) the strongly ionic limit $\tilde{E}_1 - \tilde{E}_2 \gg |t|$

³Note the sign of t depends on the symmetry of the orbitals: for s-states, with an attractive potential $V_i < 0$, then t is negative; but for p_x states t is positive for atoms aligned along x.



Figure 2.3: A simple model of a diatomic molecule. The atomic hamiltonian is $H_i = T + V_i(r)$, with T the kinetic energy $-\hbar^2 \nabla^2/2m$ and V_i the potential. We keep just one energy level on each atom.

Metals

Metals are generally characterised by a high electrical conductivity, arising because the electrons are relatively free to propagate through the solid.

Close packing. Simple metals (e.g. alkalis like Na, and s-p bonded metals such as Mg and Al) usually are highly coordinated (i.e. fcc or hcp - 12 nearest neighbours, sometimes bcc - 8 nearest neighbours), since the proximity of many neighbouring atoms facilitates hopping between neighbours. Remember that the fermi energy of a free electron gas (i.e. the average kinetic energy per particle) is proportional to $k_F^2 \propto a^{-2} \propto n^{2/3}$ (here *a* is the lattice constant and *n* the density; the average coulomb interaction of an electron in a solid with all the other electrons and the other ions is proportional to $a^{-1} \propto n^{1/3}$. Thus the higher the density, the larger the kinetic energy relative to the potential energy, and the more *itinerant* the electrons.⁴ By having a high coordination number, one can have relatively large distances between neighbours - minimising the kinetic energy cost - in comparison to a loose-packed structure of the same density.

Screening. Early schooling teaches one that a metal is an equipotential (i.e. no electric fields). We shall see later that this physics in fact extends down to scales of the screening length $\lambda \approx 0.1 nm$, i.e. about the atomic spacing (though it depends on density) - so that the effective interaction energy between two atoms in a metal is not Z^2/R (Z the charge, R the separation), but $Z^2 e^{-R/\lambda}/R$ and the cohesion is weak.

Trends across the periodic table. As an s-p shell is filled (e.g. Na,Mg,Al,Si) the ion core potential seen by the electrons grows. This makes the density of the metal tend to increase. Eventually, the preference on the right-hand side of the periodic table is for covalent semiconductor (Si, S) or insulating molecular (P, Cl) structures because the energy is lowered by making tightly bound directed bonds.

Transition metals. Transition metals and their compounds involve both the outer s-p electrons as well as inner d-electrons in the binding. The d-electrons are more localised and often are spin-polarised in the 3d shell when they have a strong atomic character (magnetism will be discussed later in the course). For 4d and 5d transition metals, the d-orbitals are more strongly overlapping from atom to atom and this produces the high binding energy of metals like W (melting point 3700 K) in comparison to alkali metals like Cs (melting point 300 K).

⁴Note the contrast to classical matter, where solids are stabilised at higher density, and gases/liquids at lower density.

2.3 Complex matter

Simple metals, semiconductors, and insulators formed of the elements or binary compounds like GaAs are only the beginning of the study of materials. Periodic solids include limitless possibilities of chemical arrangements of atoms in compounds. Materials *per se*, are not perhaps so interesting to the physicist, but the remarkable feature of condensed matter is the wealth of physical properties that can be explored through novel arrangements of atoms.

Many new materials, often with special physical properties, are discovered each year. Even for the element carbon, surely a familiar one, the fullerenes (e.g. C_{60}) and nanotubes (rolled up graphitic sheets) are recent discoveries. Transition metal oxides have been another rich source of discoveries (e.g. high temperature superconductors based on La_2CuO_4 , and ferromagnetic metals based on $LaMnO_3$). f-shell electron metals sometimes produce remarkable electronic properties, with the electrons within them behaving as if their mass is 1000 times larger than the free electron mass. Such quantum fluid ground states (metals, exotic superconductors, and superfluids) are now a rich source of research activity. The study of artificial meta-materials becins in one sense with doped semiconductors (and especially layered heterostructures grown by molecular beam epitaxy or MBE), but this subject is expanding rapidly due to an influx of new tools in nanomanipulation and biological materials.

Many materials are of course not crystalline and therefore not periodic. The physical description of *complex* and *soft* matter requires a separate course.

Glasses

If one takes a high temperature liquid (e.g. of a metal) and quenches it rapidly, one obtains a *frozen* structure that typically retains the structure of the hightemperature liquid. Melt-quenched alloys of ferromagnets are often prepared this way because it produces isotropic magnetic properties. For most materials the amorphous phase is considerably higher in energy than the crystalline, so the system has to be frozen rapidly, far from its equilibrium configuration. A few materials make glassy states readily, and the most common example is vitreous silica (SiO_2) . Crystalline forms of silica exist many of them!) and all are network structures where each Si bonds to four oxygen neighbours (approximately tetrahedrally) and each O is bonded to two Si atoms. Since the O^{2-} ion is nearly isotropic, the orientation of one tetrahedral group respect to a neighbouring group about the connecting Si - O - Si bond is not fixed, and this allows for many possible crystalline structures, but especially for the entropic stabilisation of the glass phase. Whatever the arrangement of atoms, all the electrons are used up in the bonding, so glass is indeed a good insulator. The characteristic feature of a strong or network glass is that on cooling the material becomes increasingly viscous, often following the Vogel-Fulcher law,

$$\eta \propto e^{\frac{C}{T-T_0}} \tag{2.17}$$

implying a divergence of the viscosity η at a temperature T_0 . Once η reaches about 10^{12} Pa s, it is no longer possible to follow the equilibrium behaviour. Consequently, debates still rage about whether or not the glass transition is a "true" phase transition, or indeed whether or no the temperature T_0 has physical meaning.

Polymers

The classic polymers are based on carbon, relying on its remarkable ability to adopt a variety of local chemical configurations. Polyethylene is built from repeating units of CH_2 , and more complex polymers are constructed out of more complex subunits. Because the chains are long, and easily deformed or entangled, most polymers are glassy in character, and therefore their physical properties are largely dominated by entropic considerations. The elasticity of rubber is produced by the decrease in entropy upon stretching, not by the energetic cost of stretching the atomic bonds. Many simple polymers are naturally insulating (e.g. the alkanes) or semiconducting, but it is sometimes possible to "dope" these systems so that there are electronic states. They become interesting for a number of reasons in technology and fundamental science. Because a simple polymer chain can often be modelled as a one-dimensional wire, they provide a laboratory for the often unusual properties of one-dimensional electronic systems. Because the tools of organic chemistry allow one to modify the physical properties of polymers in a wide range of ways (for example, by adding different side chains to the backbones), one can attempt to tune the electronic and optic properties of heterogeneous polymer structures to make complex devices (solar cells, light-emitting diodes, transistors) using a very different medium from inorganic semiconductors.

Liquid crystals

Like a single atom, polymers are isotropic, because they are very long. Shorter rod-shaped molecules however have an obvious orientational axis, and when combined together to make a liquid crystal one can construct matter whose properties are intermediate between liquid and solid.

Nematics. An array of rods whose centres are arranged randomly has no long-range positional order (just like a liquid), but if the rods are oriented parallel to each other has *long-range orientational order*, like a molecular crystal. This is a nematic liquid crystal. The direction in space of the orientational order is a vector \hat{n} called the *director*. The refractive index of the material will now be different for light polarized parallel and perpendicular to the director.

Cholesterics. It turns out if the molecule is *chiral* then the director need not point always in the same direction, and in a cholesteric liquid crystal the direction of \hat{n} twists slowly in a helix along an axis that is perpendicular to it. Usually the pitch of the twist is much longer than size of the rod, is a strong

function of temperature, and frequently close to the wavelength of visible light.

Smectics. Smectics additionally have long-range positional order along one direction, usually to be thought of as having layers of molecules. So called *Smectic A* has the director parallel to the planes, whereas in *Smectic C* the director is no longer perpendicular (and may indeed rotate as a function of position). In *Smectic B* the molecules in the plane have a crystalline arrangement, but different layers fall out of registry. This is a kind of quasi-2D solid.





Schematic representation of the position and orientation of anisotropic molecules in: (a) the isotropic phase; (b) the nematic phase; (c) the smectic-A phase; and (d) the smectic-C phase. [From Chaikin and Lubensky]

2.3. COMPLEX MATTER

Quasicrystals

As a last piece of exotica, the classic group theory of crystal structures proves the impossibility of building a Bravais lattice with five-fold symmetry. Nature is unaware of this, and a series of metallic alloys have been found that indeed have crystals with axes of three, five, and ten-fold symmetry. These materials are in fact physical representations of a mathematical problem introduced by Penrose of tiling of a plane with (e.g.) two rhombus shaped tiles that have corner angles of $2\pi/10$ and $2\pi/5$. A complete tiling of the plane is possible, though the structure is not a periodic lattice (it never repeats).



Figure 2.5: Scanning tunnelling microscope image of a $10 \ nm^2$ quasicrystal of AlPdMn with a Penrose tiling overlaid. [Ledieu et al Phys.Rev.B **66**, 184207 (2002)]

Chapter 3

The Fermi and Bose gases

Most of this is revision material.

3.1 Free electron gas in three-dimensions

Consider a free electron gas, confined to a three-dimensional box of side L. The free particle Schrodinger equation is

$$-\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial y^2} \right) \psi(\mathbf{r}) = \epsilon \psi(\mathbf{r})$$
(3.1)

which has the following eigenstates:

$$\psi_{\mathbf{k}}(\mathbf{r}) = \mathcal{N}\sin(k_x x)\sin(k_y y)\sin(k_z z) \tag{3.2}$$

with energy

$$\epsilon_{\mathbf{k}} = \frac{\hbar^2 |k|^2}{2m} \tag{3.3}$$

Owing to the restriction to the box (0 < x < L, 0 < y < L, 0 < z < L)) the allowed values of k are discrete.

$$\mathbf{k} = \frac{\pi}{L}(n_x, n_y, n_z) \tag{3.4}$$

with n_x , n_y , n_z positive integers.

It is more convenient to introduce wavefunctions that satisfy periodic boundary conditions, namely

$$\psi(x+L, y, z) = \psi(x, y, z) \tag{3.5}$$

and similarly for y and z directions. These are of the form of a plane wave

$$\psi_k(\mathbf{r}) = \exp(i\mathbf{k} \cdot \mathbf{r}) \tag{3.6}$$

where the eigen-energies are identical to Eq. (3.4) but the restriction on momentum being

$$\mathbf{k} = \frac{2\pi}{L}(n_x, n_y, n_z) \tag{3.7}$$

with n_x , n_y , n_z positive or negative integers.

Together with the spin quantum number m, the components of k are the good quantum numbers of the problem.

Qu.3.3 Momentum operator

Show that the state $\psi_k(\mathbf{r}) = \exp(i\mathbf{k}\cdot\mathbf{r})$ is an eigenstate of the momentum operator $\hat{\mathbf{p}} = -i\hbar\vec{\nabla}$ and find the eigenvalue.

3.2 Fermi surface, and density of states

In the ground state at zero temperature, the fermi gas can then be represented by filling up all the low energy states up to a maximum energy ϵ_F (the fermi energy) corresponding to a sphere of radius the fermi momentum k_F in k-space.

Each triplet of quantum numbers k_x , k_y , k_z accounts for two states (spin degeneracy) and occupies a volume $(2\pi/L)^3$.

The total number of occupied states inside the fermi sphere is

$$N = 2 \cdot \frac{4/3\pi k_f^3}{(2\pi/L)^3} \tag{3.8}$$

so that the fermi wave-vector is written in term of the electron density n = N/V as

$$k_f = (3\pi^2 n)^{1/3} \tag{3.9}$$

1

We are often interested in the *density of states*, g(E), which is the number of states per unit energy range. Calculate it by determining how many states are enclosed by a thin shell of energy width dE, viz.

$$g(E)dE = 2 \cdot \frac{\text{Volume of shell in } k - \text{space}}{\text{Volume of } k - \text{space per state}} = 2 \cdot \frac{4\pi k^2 dk}{(2\pi)^3/V} \quad , \tag{3.10}$$

hence

$$g(E) = 2\frac{V}{(2\pi)^3} 4\pi k^2 \frac{dk}{dE} = \frac{V}{\pi^2} \frac{m}{\hbar^2} \left(\frac{2mE}{\hbar^2}\right)^{\frac{1}{2}} \quad . \tag{3.11}$$

The factor of 2 is for spin degeneracy. Often, the density of states is given per unit volume, so the factor of V disappears.

26

Qu.3.4 Density of states for free electrons

(a) What is the fermi wavevector and fermi energy as a function of particle density for a free electron gas in one and two dimensions (define density appropriately)?

(b) Calculate the density of states in energy for free electrons in one and two dimensions. [Answer: $(2m/\pi\hbar^2) \times (\hbar^2/2mE)^{\frac{1}{2}}$, (d=1); $(m/\pi\hbar^2)$, d=2; $(m/\pi^2\hbar^2) \times (2mE/\hbar^2)^{\frac{1}{2}}$, d=3 .]

(c) Show how the 3D density of states can be re-written as

$$(3/2)(n/E_F)(E/E_F)^{\frac{1}{2}}$$

with n = N/V.

3.3 Thermal properties of the electron gas

The occupancy of states in thermal equilibrium in a fermi system is governed by the fermi distribution

$$f(E) = \frac{1}{e^{(E-\mu)/k_B T} + 1}$$
(3.12)

where the chemical potential μ can be identified (at zero temperature) with the fermi energy E_F of the previous section.

The number density of particles is

$$n = N/V = \frac{1}{V} \sum_{i} f(E_{i}) = \frac{2}{V} \sum_{\mathbf{k}} f(\epsilon_{\mathbf{k}})$$
$$= \frac{1}{4\pi^{3}} \int d\mathbf{k} f(\epsilon_{\mathbf{k}})$$
$$= \int dE \ g(E) f(E)$$
(3.13)

The internal energy density u = U/V can then be written in the same fashion:

$$u = \int dE \ Eg(E)f(E) \tag{3.14}$$

Eq. (3.14) will be used to derive the electronic specific heat $c_v = \partial u / \partial T |_v$ at constant volume. The estimation is made much simpler by realising that in almost all cases of interest, the energy scale set by temperature $k_B T$ (≈ 0.025 eV at room temperature) is much less than the fermi energy E_F (a few eV in most metals). From Eq. (3.14)

$$c_v = \int dE \ Eg(E) \frac{\partial f(E)}{\partial T}$$
(3.15)

Notice that the fermi function is very nearly a step-function, so that the temperaturederivative is a function that is sharply-peaked for energies near the chemical potential. The contribution to the specific heat then comes only from states within k_BT of the chemical potential and is much less than the $3/2k_B$ per particle from classical distinguishable particles. From such an argument, one guesses that the specific heat per unit volume is of order

$$c_v \approx \frac{N}{V} \frac{k_B T}{E_F} k_B \tag{3.16}$$

Doing the algebra is a little tricky, because it is important to keep the number density to stay fixed (Eq. (3.13)) — which requires the chemical potential to shift (a little) with temperature since the density of states is not constant. A careful calculation is given by Ashcroft and Mermin.

But to the extent that we can take the density of states to be a constant, we can remove the factors g(E) from inside the integrals. Notice that with the change of variable $x = (E - \mu)/k_BT$

$$\frac{df}{dT} = \frac{e^x}{(e^x + 1)^2} \times \left[\frac{x}{T} + \frac{1}{k_B T} \frac{d\mu}{dT}\right]$$
(3.17)

The number of particles is conserved, so we can write

$$\frac{dn}{dT} = 0 = g(E_F) \int dE \ \frac{\partial f(E)}{\partial T}$$
(3.18)

which on using Eq. (3.17) becomes

$$0 = g(E_F)k_BT \int_{-\infty}^{\infty} dx \frac{e^x}{(e^x + 1)^2} \times \left[\frac{x}{T} + \frac{1}{k_BT}\frac{d\mu}{dT}\right] \quad . \tag{3.19}$$

The limits can be safely extended to infinity: the factor $\frac{e^x}{(e^x+1)^2}$ is even, and hence at this level of approximation $d\mu/dT = 0$.

To the same level of accuracy, we have

$$c_v = g(E_F) \int dE \ E \frac{\partial f(E)}{\partial T}$$

= $g(E_F)k_B T \int_{-\infty}^{\infty} dx \ (\mu + k_B T x) \frac{e^x}{(e^x + 1)^2} \frac{x}{T}$
= $g(E_F)k_B^2 T \int_{-\infty}^{\infty} dx \ \frac{x^2 e^x}{(e^x + 1)^2}$
= $\frac{\pi^2}{3} k_B^2 T g(E_F)$ (3.20)

The last result is best understood when rewritten as

$$c_v = \frac{\pi^2}{2} \frac{k_B T}{E_F} n k_B \tag{3.21}$$

confirming the simple argument given earlier and providing a numerical prefactor.

28

3.4. LATTICE DYNAMICS AND PHONONS

The calculation given here is just the leading order of an expansion in powers of $(k_BT/E_F)^2$. To next order, one finds that the chemical potential is indeed temperature-dependent:

$$\mu = E_F \left[1 - \frac{1}{3} \left(\frac{\pi k_B T}{2E_F} \right)^2 + O(k_B T / E_F)^4 \right]$$
(3.22)

but this shift is small in a dense metal at room temperature, and may usually be neglected.

Qu.3.5 Thermodynamic properties of a free electron metal

Derive the free electron formula for the fermi energy E_F , the fermi wavevector k_F and the density of states at the fermi level $g(E_F)$.

Within the free electron model at zero temperature:

Show that the total energy for N electrons is $\overline{E} = \frac{3}{5}NE_F$.

Calculate the pressure, p, using $p = -\frac{d\bar{E}}{d\Omega}$, where Ω is the volume.

Calculate the bulk modulus $B = -\Omega \frac{dp}{d\Omega}$.

Potassium is monovalent and has an atomic concentration of $1.402\times10^{28}m^{-3}.$ Compare the bulk modulus calculated above with the experimental value of 3.7×10^9 Pa.

Estimate $g(E_F)$ for magnesium, which has a valence of 2 and an atomic concentration of $4.3\times 10^{28}~m^{-3}$. Use this value to estimate the asymptotic low temperature specific heat, compared to the experimental value of $c_v/T=1.3\,mJ\,mol^{-1}\,K^{-2}$.

3.4 Lattice dynamics and phonons

One-dimensional monatomic chain

Our model consists of identical atoms connected by springs, shown in Fig. 3.1

In equilibrium, the atoms are uniformly spaced at a distance a, and we now look for oscillations about the equilibrium position. We assume the crystal is harmonic, so that the spring restoring force is linearly dependent upon the extension. Then if we take the displacement of the n^{th} atom (which is at the point $r_n = na$) to be u_n , its equation of motion is

$$m\frac{\partial^2 u_n}{\partial t^2} = K(u_{n+1} - u_n) + K(u_{n-1} - u_n)$$
(3.23)

We guess that the solution is a wave, of the form

$$u_n(t) = u_o \cos(qr_n - \omega(q)t) \tag{3.24}$$



Figure 3.1: A one-dimensional linear chain. The atoms are shown in their equally spaced equilibrium conditions in the top row, and with a periodic distortion below. The bottom figure plots the displacements u_n as arrows, and the curve shows how this is a sine-wave of period 6a, in this case.

Here the wavelength of the wave is $\lambda = 2\pi/q$, and the period is $T = 2\pi/\omega(q)$; to check that this is a solution, and to determine the frequency we substitute in the equation of motion. To do this is left as an exercise, and a few lines of algebra will show that the solution Eq. (3.24) exists provided that

$$m\omega^2(q) = 2K(1 - \cos(qa)) = 4K\sin^2(\frac{qa}{2})$$
 (3.25)

so that

$$\omega(q) = 2(K/m)^{1/2} \sin(\frac{qa}{2})$$
(3.26)

Eq. (3.25) is called a dispersion relation — the relation between the frequency of the mode and its wavevector, or equivalently the relationship between the wavelength and the period.

Qu.3.6 Acoustic phonon dispersion in the monatomic chain By substituting Eq. (3.24) in Eq. (3.23) derive the dispersion relation Eq. (3.25) for the one-dimensional monatomic chain.

The wavevector q is inversely related to the wavelength; note that for long wavelength modes (i.e. $q \rightarrow 0$), the relationship is linear, viz

$$\omega(q) = (K/m)^{1/2}(qa) \tag{3.27}$$

which is the same as for a wire with tension Ka and density m/a. In the long wavelength limit, we have compressive sound waves that travel with a velocity $v = a(K/m)^{1/2}$. Because this kind of wave behaves like a sound wave, it is generally called an acoustic mode.



Figure 3.2: Dispersion relation between frequency and wavevector for a onedimensional monatomic chain

The dispersion is not linear for larger values of q, and is in fact periodic (Fig. 3.2). The periodicity can easily be understood by reference to Eq. (3.24). Suppose we choose $q = 2\pi/a$. Note then that

$$qr_n = \frac{2\pi}{a} \times na = 2\pi n \tag{3.28}$$

so that all the atoms displace together, just as if q = 0. In general it is straightforward to show that if one replaces q by q+integer $\times 2\pi a$, then the displacements are unchanged – so we may simplify our discussion by using only q vectors in the range

$$-\frac{\pi}{a} \le q \le \frac{\pi}{a} \quad . \tag{3.29}$$

This is called the first Brillouin zone.

One-dimensional diatomic chain

The monatomic chain contains only acoustic modes, but the phonon spectrum becomes more complex if there are more atoms per unit cell. As an illustration, we look at the diatomic chain.

For simplicity, we use again a phenomenological model of balls and springs, but now with two different atoms in the unit cell, two different masses and two different spring constants (see Fig. 3.3). We can now write down two equations of motion, one for each type of atom:

$$m_{A} \frac{\partial^{2} u_{nA}}{\partial t^{2}} = K(u_{nB} - u_{nA}) + K'(u_{n-1,B} - u_{nA})$$
$$m_{B} \frac{\partial^{2} u_{nB}}{\partial t^{2}} = K'(u_{n+1A} - u_{nB}) + K(u_{n,A} - u_{nB})$$
(3.30)



Figure 3.3: Diatomic chain

The solution of this is a little more complicated than before, but we can now intuitively see that there ought to be a new type of phonon mode by considering a particular limit of the parameters. Suppose the two atoms are quite strongly bound together in pairs, as sketched in the figure above: then we might expect that $K \gg K'$, and to a first approximation the pairs can be treated as independent molecules. (We will also simplify the analysis by taking $m_A = m_B = m$.) Then every molecule will have a vibrational mode where the two atoms oscillate out of phase with each other with a frequency

$$\omega_o^2 = 2K/m \quad . \tag{3.31}$$

The corresponding coodinate which undergoes this oscillation is

$$u_{opt}(q=0) = u_A - u_B \tag{3.32}$$

where I have explicitly remarked that this is at q = 0 if each molecule undergoes the oscillation in phase with the next.

We can of course make a wavelike solution by choosing the correct phase relationship from one unit cell to the next — as sketched in Fig. 3.4, but if $K' \ll K$ this will hardly change the restoring force at all, and so the frequency of this so-called optical phonon mode will be almost independent of q.



Figure 3.4: Dispersion of the optical and acoustic phonon branches in a diatomic chain, and a schematic picture of the atomic displacements in the optical mode at q=0

There are now two branches of the dispersion curve, along one of which the frequency vanishes linearly with wavevector, and where the other mode has a

finite frequency as $q \rightarrow 0$ (see Fig. 3.5). The name "optical" arises because at these long wavelengths the optical phonons can interact (either by absorption, or scattering) with light, and are therefore prominent features in the absorption and Raman spectra of solids in the infrared spectrum.



Figure 3.5: Pattern of atomic displacements for an acoustic and an optical phonon of the same wavevector.

Qu.3.7 * Acoustic and optic phonons in the diatomic chain

This question involves somewhat messy algebra to derive the dispersion relation for the diatomic chain.

In the diatomic chain, we take the unit cell to be of length a, and take x_A and x_B to be the coordinates of the A and B atoms within the unit cell. Hence, in the n^{th} cell,

$$r_{n,A} = na + x_A; \quad r_{n,B} = na + x_B \quad .$$
 (3.33)

In the equations of motionEq. (3.30), look for solutions of the form

$$u_{n,\alpha} = e_{\alpha}(q) \exp i(qr_{n,\alpha} - \omega(q)t) + e_{\alpha}^{*}(q) \exp i(-qr_{n,\alpha} + \omega(q)t)$$
(3.34)

where $\alpha = A$ or B, and e_{α} are complex numbers that give the amplitude and phase of the oscillation of the two atoms.

Separating out the terms that have the same time dependence, show that (for equal masses, $m_A = m_B = m$)

$$m\omega^{2}(q)e_{A}(q) = D_{AA}(q)e_{A}(q) + D_{AB}(q)e_{B}(q)$$

$$m\omega^{2}(q)e_{B}(q) = D_{BA}(q)e_{A}(q) + D_{BB}(q)e_{B}(q)$$
(3.35)

where

$$D_{AA}(q) = D_{BB}(q) = K + K' \quad , \tag{3.36}$$

$$-D_{AB}(q) = K \exp iq(r_{n,B} - r_{n,A}) + K' \exp iq(r_{n-1,B} - r_{n,A}) -D_{BA}(q) = K \exp iq(r_{n,A} - r_{n,B}) + K' \exp iq(r_{n+1,A} - r_{n,B})$$
(3.37)

Check that $D_{AB} = D_{BA}^*$.

The 2x2 matrix equation can have a non-trivial solution if the determinant vanishes:

$$\begin{vmatrix} D_{AA}(q) - m\omega^2(q) & D_{AB}(q) \\ D_{BA}(q) & D_{BB}(q) - m\omega^2(q) \end{vmatrix} = 0$$
(3.38)

Hence show that the frequencies of the modes are given by

$$m\omega^2(q) = K + K' \pm \left[(K + K')^2 - 2KK' \sin^2(\frac{qa}{2}) \right]^{1/2} \quad . \tag{3.39}$$

Sketch the dispersion relations when K/K' = 2.

Discuss what happens if K = K'.

Phonons in three-dimensional solids

The descriptions above are not too hard to generalise to three- dimensional solids, although the algebra gets overloaded with suffices.

Rather than a one-dimensional wavevector k corresponding to the direction of the 1D chain, there is now a three-dimensional dispersion relation $\omega(\vec{k})$, describing waves propagating in different directions.

Also, there are not just compressional waves, but also transverse, or shear waves, that will have different dispersion from the longitudinal (compressional) waves. (These exist in a crystal in any dimension, including our 1D chain, where they can be imagined to involve displacements perpendicular to the chain direction.) Quite generally, for each atom in the unit cell, one expects to find three branches of phonons (two transverse, and one longitudinal); always there are three acoustic branches, so a solid that has m atoms in its unit cell will have 3(m-1) optical modes. And again, each optical modes will be separated into two transverse branches and one longitudinal branch.¹

Density of states

Just as for the electron gas problem we need to write down the density of states for phonons. First, we need to count how many modes we have and understand their distribution in momentum space.

In the 1D monatomic chain containing N atoms (assume N very large), there are just N degrees of freedom (for the longitudinal vibration) and therefore N

¹The separation between longitudinal and transverse is only rigorously true along lines of symmetry in \vec{k} -space.

3.4. LATTICE DYNAMICS AND PHONONS

modes. This tells us (and we can see explicitly by looking at boundary conditions for an N-particle chain) that the allowed k-points are discrete, viz

$$k_n = \frac{2\pi}{L}n \; ; \; n = \left(-\frac{N}{2}, -\frac{N-1}{2}, ..., \frac{N}{2}\right] \; ,$$
 (3.40)

so that k runs from $-\pi/a$ to π/a , with a = N/L, the lattice constant. Notice this is the same spacing of k-states for the electron problem, and the only difference is that because the atoms are discrete, there is a maximum momentum (on the Brillouin zone boundary) allowed by counting degrees of freedom.

By extension, in three dimensions, each branch of the phonon spectrum still contains N states in total, but now $N = L^3/\Omega_{cell}$ with Ω_{cell} the volume of the unit cell, and $L^3 = V$ the volume of the crystal. The volume associated with each allowed k-point is then

$$\Delta k = \frac{(2\pi)^3}{L^3} \tag{3.41}$$

There are 3 acoustic branches, and 3(m-1) optical branches.

It is convenient to start with a simple description of the optical branch(es), the *Einstein* model, which approximates the branch as having a completely flat dispersion $\omega(\mathbf{k}) = \omega_0$. In that case, the density of states in frequency is simply

$$D_E(\omega) = N\delta(\omega - \omega_0) \quad . \tag{3.42}$$

We have a different results for the acoustic modes, which disperse linearly with momentum as $\omega \to 0$. Using a dispersion $\omega = vk$, and following the earlier argument used for electrons, we get the *Debye* model

$$D_D(\omega) = \frac{4\pi k^2}{(2\pi/L)^3} \frac{dk}{d\omega} = \frac{V\omega^2}{2\pi v^3} .$$
 (3.43)

Of course this result cannot apply once the dispersion curves towards, the zone boundary, and there must be an upper limit to the spectrum. In the Debye model, we cut off the spectrum at a frequency ω_D , which is determined so that the total number of states (N) is correctly counted, i.e. by choosing

$$\int_{0}^{\omega_{D}} d\omega D_{D}(\omega) = N \tag{3.44}$$

which yields

$$\omega_D^3 = \frac{6\pi^2 v^3 N}{V} \quad . \tag{3.45}$$

Notice that this corresponds to replacing the correct cutoff in momentum space (determined by intersecting Brillouin zone planes) with a sphere of radius

$$k_D = \omega_D / v \quad . \tag{3.46}$$



Figure 3.6: Comparison of Debye density of states (a) with that of a real material (b).

3.5 Lattice specific heat

Phonons obey Bose-Einstein statistics, but their number is not conserved and so the chemical potential is zero, leading to the Planck distribution

$$n(\omega) = \frac{1}{\exp(\hbar\omega/k_B T) - 1} \quad . \tag{3.47}$$

The internal energy is

$$U = \int d\omega D(\omega) n(\omega) \hbar \omega$$
 (3.48)

For the Einstein model

$$U_E = \frac{N\hbar\omega_o}{e^{\hbar\omega_o/k_BT} - 1} \tag{3.49}$$

and the heat capacity

$$C_V = \left(\frac{\partial U}{\partial T}\right)_V = Nk_b \left(\frac{\hbar\omega_o}{k_B T}\right)^2 \frac{e^{\hbar\omega_o/k_B T}}{(e^{\hbar\omega_o/k_B T} - 1)^2} \quad . \tag{3.50}$$

At low temperatures, this grows as $exp - \hbar\omega_o/k_BT$ and is very small, but it saturates at a value of Nk_B (the Dulong and Petit law) above the characteristic temperature $\theta_E = \hbar\omega_o/k_B$.²

At low temperature, the contribution of optical modes is small, and the Debye spectrum is appropriate. This gives

$$U_D = \int_0^{\omega_D} d\omega \frac{V\omega^2}{2\pi^2 v^3} \frac{\hbar\omega}{e^{\hbar\omega/k_B T} - 1} \quad . \tag{3.51}$$

 $^{^2\}mathrm{This}$ is per branch of the spectrum, so gets multiplied by 3 in three dimensions
3.5. LATTICE SPECIFIC HEAT

Power counting shows that the internal energy then scales with temperature as T^4 and the specific heat as T^3 at low temperatures. The explicit formula can be obtained as

$$C_V = 9Nk_B \left(\frac{T}{\theta_D}\right)^3 \int_0^{\theta_D/T} dx \frac{x^4 e^x}{(e^x - 1)^2} , \qquad (3.52)$$

where the *Debye temperature* is $\theta_D = \hbar \omega / k_B$. We have multiplied by 3 to account for the three acoustic branches.

Qu.3.8 Lattice specific heat

From Eq. (3.51) derive the formula for the Debye specific heat Eq. (3.52).

Evaluate the integral at *high* temperature $T \gg \theta_D$, and therefore determine the high temperature behaviour of the specific heat.

Using the formula $\int_0^\infty \, dx \frac{x^4 e^x}{(e^x-1)^2} = \frac{4\pi^4}{15}$, determine the low temperature behaviour of the Debye specific heat.

Sketch the heat capacity formulae from the Debye and Einstein models and compare them.

Chapter 4

Periodic solids and diffraction

4.1 The description of periodic solids

An ideal crystal is constructed from the infinite repetitition of identical structural units in space. The repeating structure is called the *lattice*, and the group of atoms which is repeated is called the *basis*. The basis may be as simple as a single atom, or as complicated as a polymer or protein molecule. This section discusses briefly some important definitions and concepts. For a more complete description with examples, see any of the textbooks recommended in the introduction.

Lattice. The lattice is defined by three fundamental (called *primitive*) translation vectors \mathbf{a}_i , i = 1, 2, 3. The atomic arrangement looks the same from equivalent points in the unit cell:

$$\mathbf{r}' = \mathbf{r} + \sum_{i} n_i \mathbf{a}_i \quad \forall \text{ integer } n_i .$$
 (4.1)

Primitive unit cell. The primitive unit cell is the parallelipiped formed by the primitive translation vectors \mathbf{a}_i , and an arbitrary lattice translation operation can be written as

$$\mathbf{T} = \sum_{i} n_i \mathbf{a}_i \tag{4.2}$$

There are many ways of choosing a primitive unit cell, but the lattice so formed is called a *Bravais* lattice.

Wigner-Seitz cell A most convenient primitive unit cell to use is the *Wigner-Seitz* cell, constructed as follows: Draw lines to connect a given lattice point to all of its near neighbours. Then draw planes normal to each of



Figure 4.1: . The Wigner-Seitz cell for the BCC and FCC lattices

these lines from the midpoints of the lines. The smallest volume enclosed in this way is the Wigner-Seitz primitive unit cell.

Point group. The are other symmetry operations that can be performed on a lattice, for example rotations and reflections. The collection of symmetry operations, which applied about a lattice point, map the lattice onto itself is the lattice *point group*. This includes reflections and rotations; for example a 2D square lattice is invariant under reflections about the x and y axes, as well as through axes at an angle of $\pi/4$ to the x and y axes, and rotations through any multiple of $\pi/2$. Remember that adding a basis to a primitive lattice may destroy some of the point group symmetry operations. There are five distinct lattice types in two dimensions, and 14 in three dimensions.

Space group. The translational symmetry and the point group symmetries are subgroups of the full symmetry of the lattice which is the *space* group. Every operation in the space group consists of a rotation, reflection, or inversion followed by a translation. However, the space group is not necessarily just the sum of the translational symmetries and the point symmetries, because there can be space group symmetries that are the sum of a proper rotation and a translation, neither of which are independently symmetries of the lattice.

The number of possible lattices is large. In three dimensions there are 32 distinct point groups, and 230 possible lattices with bases. Two of the important lattices that we shall meet later are the body-centred and face-centred cubic lattices, shown in Fig. 4.1.

Index system for crystal planes

If you know the coordinates of three points (not collinear), this defines a plane. Suppose you chose each point to lie along a different crystal axis, the plane is then specified by giving the coordinates of the points as

$$x\mathbf{a}_1 + y\mathbf{a}_2 + z\mathbf{a}_3 \tag{4.3}$$



Figure 4.2: Illustration of Bragg scattering from a crystal

Of course the triad (xyz) need not be integers. However, one can always find a plane parallel to this one by finding a set of three integers (hkl) where x/h = y/k = z/l. (hkl) is called the *index* of the plane. When we want to refer to a set of planes that are equivalent by symmetry, we will use a notation of curly brackets: so $\{100\}$ for a cubic crystal denotes the six equivalent symmetry planes (100), (010), (001), $(\overline{1}00)$, $(0\overline{1}0)$, $(00\overline{1})$, with the overbar used to denote negation.

4.2 The reciprocal lattice and diffraction

The reciprocal lattice as a concept arises from the theory of the scattering of waves by crystals. You should be familiar with the diffraction of light by a 2-dimensional periodic object - a diffraction grating. Here an incident plane wave is diffracted into a set of different directions in a Fraunhofer pattern. An infinite periodic structure produces outgoing waves at particular angles, which are determined by the periodicity of the grating. What we discuss now is the generalisation to scattering by a three-dimensional periodic lattice.

First calculate the scattering of a single atom (or more generally the basis that forms the unit cell) by an incoming plane wave, which should be familiar from elementary quantum mechanics. An incoming plane wave of wavevector \mathbf{k}_o is incident on a potential centred at the point \mathbf{R} . At large distances the scattered wave take the form of a circular wave. (See figure Fig. 4.2) The total field (here taken as a scalar) is then

$$\psi \propto e^{i\mathbf{k}_o \cdot (\mathbf{r} - \mathbf{R})} + cf(\hat{r}) \frac{e^{ik_o|\mathbf{r} - \mathbf{R}|}}{|\mathbf{r} - \mathbf{R}|}$$
(4.4)

All the details of the scattering is in the form factor $f(\hat{r})$ which is a function of the scattering angle, the arrangement and type of atom, etc. The total scattered intensity is just set by c and we will assume it is small (for this reason we do not consider multiple scattering by the crystal)

For sufficiently large distance from the scatterer, we can write

$$k_o |\mathbf{r} - \mathbf{R}| \approx k_o r - k_o \frac{\mathbf{r} \cdot \mathbf{R}}{r}$$
(4.5)

Define the scattered wavevector

$$\mathbf{k} = k_o \frac{\mathbf{r}}{r} \tag{4.6}$$

and the momentum transfer

$$\mathbf{q} = \mathbf{k}_o - \mathbf{k} \tag{4.7}$$

we then have for the waveform

$$\psi \propto e^{i\mathbf{k}_o \cdot \mathbf{r}} \left[1 + cf(\hat{r}) \frac{e^{i\mathbf{q}\cdot\mathbf{R}}}{r} \right]$$
 (4.8)

Now sum over all the identical sites in the lattice, and the final formula is

$$\psi \propto e^{i\mathbf{k}_o \cdot \mathbf{r}} \left[1 + c \sum_i f_i(\hat{r}) \frac{e^{i\mathbf{q} \cdot \mathbf{R}_i}}{r} \right] \quad . \tag{4.9}$$

Away from the forward scattering direction, the incoming beam does not contribute, and we need only look at the summation term. We are adding together terms with different phases $\mathbf{q} \cdot \mathbf{R}_i$, and these will lead to a cancellation unless the Bragg condition is satisfied

$$\mathbf{q} \cdot \mathbf{R} = 2\pi m \tag{4.10}$$

for all **R** in the lattice, and with *m* an integer (that depends on **R**). The special values of $\mathbf{q} \equiv \mathbf{G}$ that satisfy this requirement lie on a lattice, which is called the *reciprocal lattice*.¹

One can check that the following prescription for the reciprocal lattice will satisfy the Bragg condition. The primitive vectors \mathbf{b}_i of the reciprocal lattice are given by

$$\mathbf{b}_1 = 2\pi \frac{\mathbf{a}_2 \wedge \mathbf{a}_3}{\mathbf{a}_1 \cdot \mathbf{a}_2 \wedge \mathbf{a}_3} \quad \text{and cyclic permutations} \quad . \tag{4.11}$$

Qu.4.9 BCC and FCC lattices Show that the reciprocal lattice of a body centred cubic lattice (BCC) of spacing a is a face centred cubic (FCC) lattice of spacing $4\pi/a$; and that the reciprocal lattice of a FCC lattice of spacing a is a BCC lattice of spacing $4\pi/a$.

 $^{^{1}}$ We can be sure that they are on a lattice, because if we have found any two vectors that satisfy Eq. (4.10), then their sum also satisfies the Bragg condition.

Qu.4.10 Reciprocal lattice cell volume Show that the volume of the primitive unit cell of the reciprocal lattice is $(2\pi)^3/\Omega_{cell}$, where Ω_{cell} is the volume of the primitive unit cell of the crystal.

4.3 Diffraction conditions and Brillouin zones

For elastic scattering, there are two conditions relating incident and outgoing momenta. Conservation of energy requires that the magnitudes of k_o and k are equal, and the Bragg condition requires their difference to be a reciprocal lattice vector $\mathbf{k} - \mathbf{k}_o = \mathbf{G}$. The combination of the two can be rewritten as

$$\mathbf{k} \cdot \frac{\mathbf{G}}{2} = (\frac{G}{2})^2 \quad . \tag{4.12}$$

Eq. (4.12) defines a plane constructed perpendicular to the vector **G** and intersecting this vector at its midpoint. The set of all such planes defines those incident wavevectors that satisfy the conditions for diffraction (see Fig. 4.3).



Figure 4.3: Ewald construction. The points are the reciprocal lattice of the crystal. k_0 is the incident wavevector, with the origin chosen so that it terminates on a reciprocal lattice point. A sphere of radius $|k_0|$ is drawn about the origin, and a diffracted beam will be formed if this sphere intersects any other point in the reciprocal lattice. The angle θ is the Bragg angle of Eq. (4.13)

This condition is familiar as Bragg's Law.

Qu.4.11 Bragg's law (a) Show that the reciprocal lattice vector $\mathbf{G} = h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_2$ is perpendicular to the (hkl) plane of the crystal lattice.

(b) Show that the distance between two adjacent (hkl) planes is $2\pi/|\mathbf{G}|$.

(c) Show that the condition Eq. (4.12) may be written as

$$\frac{2\pi}{\lambda}\sin\theta = \frac{\pi}{d}\tag{4.13}$$

where $\lambda = 2\pi/k$, and θ is the angle between the incident beam and the crystal plane.

Since the indices that define an actual crystal plane may contain a common factor n, whereas the definition used earlier for a *set* of planes removed it, we should generalise Eq. (4.13) to define d to be the spacing between adjacent parallel planes with indices h/n, k/n, l/n. Then we have

$$2d\sin\theta = n\lambda\tag{4.14}$$

which is the conventional statement of Bragg's Law.

To recap:

- The set of planes that satisfy the Bragg condition can be constructed by finding those planes that are perpendicular bisectors of every reciprocal lattice vector \mathbf{G} . A wave whose wavevector drawn from the origin terminates in any of these planes satisifies the condition for elastic diffraction.
- The planes divide reciprocal space up into cells. The one closest to the origin is called the first Brillouin zone. The nth Brillouin zone consists of all the fragments exterior to the $(n-1)^{th}$ plane (measured from the origin) but interior to the n^{th} plane.
- The first Brillouin zone is the Wigner-Seitz cell of the reciprocal lattice. This will play an important role in the discussion of electronic states in a periodic potential.
- The volume of each Brillouin zone (adding up the fragments) is equal to the volume of the primitive unit cell of the reciprocal lattice, which is $(2\pi)^3/\Omega_{cell}$ where Ω_{cell} is the volume of the primitive unit cell of the crystal.

Chapter 5

Electronic structure from local orbitals

5.1 Tight binding: Linear combination of atomic orbitals

Perhaps the most natural view of a solid is to think about it as a collection of interacting atoms, and to build up the wavefunctions in the solid from the wavefunctions of the individual atoms. This is the *linear combination of atomic orbital* (LCAO) or *tight-binding* method.

5.2 Diatomic molecule revisited

Remember the model problem question 2.2 where we worked with a highly restricted basis on one orbital per atom. For identical atoms, the full Hamiltonian consists of

$$H = H_1 + H_2 = T + V_1 + V_2 \tag{5.1}$$

with T the kinetic energy and V_1 , V_2 the (identical potentials) on the two atoms. The basis set consists of two states $|1\rangle$ and $|2\rangle$ that satisfy

$$H_1|1> = E_0|1>$$
 (5.2)

$$H_2|2> = E_0|2>$$
 (5.3)

so that E_0 is the eigenenergy of the atomic state, and we look for solutions

$$|\psi\rangle = c_1|1\rangle + c_2|2\rangle \tag{5.4}$$

We solve this in the usual way: Project $H|\psi\rangle = E|\psi\rangle$ onto <1| and <2

to get the simultaneous equations

$$\begin{pmatrix} \tilde{E}_0 - E & t \\ t^* & \tilde{E}_1 - E \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = 0$$
(5.5)

neglecting the overlap elements < 1|2 >.

Here

$$\tilde{E}_0 = H_{11} = \langle 1 | T + V_1 + V_2 | 1 \rangle = E_1 + \langle 1 | V_2 | \rangle$$
(5.6)

is a shift of the atomic energy by the *crystal field* of the other atom(s). The more interesting term is the *hopping* matrix element that couples the atomic states together:¹

$$t = H_{12} = \langle 1 | T + V_1 + V_2 | 2 \rangle \tag{5.7}$$

For t < 0, the new eigenstates are

$$|\psi\rangle = \frac{1}{\sqrt{2}} \left[|1\rangle \mp |2\rangle\right] \quad E = \tilde{E}_1 \pm |t| \tag{5.8}$$

The lower energy (bonding) state has electron density higher between atoms. The higher energy (antibonding) state has node between atoms.

5.3Linear chain

,

Now let us generalised this to a ring of N identical atoms (i.e. a chain with periodic boundary conditions:

$$|\psi\rangle = \sum_{i} c_i |i\rangle \tag{5.9}$$

、

、

which now generates the matrix equations²

$$\begin{pmatrix} E_0 - E & t & 0 & \dots & 0 & t \\ t & E_0 - E & t & \dots & 0 & 0 \\ 0 & t & E_0 - E & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & E_0 - E & t \\ t & 0 & 0 & \dots & t & E_0 - E \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ \dots \\ c_{N-1} \\ c_N \end{pmatrix}$$
(5.10)

The solutions are (check by substitution)

$$c_j^{(m)} = \frac{1}{\sqrt{N}} \exp\left(2\pi i \frac{jm}{N}\right) \tag{5.11}$$

$$E^{(m)} = E_o + 2t \cos\left(\frac{2\pi m}{N}\right) \quad m = 0, 1, ..., N - 1$$
 (5.12)

46

¹Note the sign of t depends on the symmetry of the orbitals: for s-states, with an attractive

potential $V_i < 0$, then t is negative; but for p_x states t is positive for atoms aligned along x. ²The t's in the corners make this matrix symmetric under translations: it is called a circulant



Figure 5.1: Eigenvalues of the 1D chain Eq. (5.11) are confined to a band in energy centred on the (shifted) atomic energy level \tilde{E}_0 . If N is very large, the energies form a continuous band and are periodic in m. Then we replace the index m by the continuous crystal momentum $k = 2\pi m/Na$, with a the lattice constant. So we could label the states more symmetrically by keeping a range -N/2 + 1 < m < N/2 (or $-\pi/a < k < \pi/a$); this is called the first Brillouin zone.

5.4 Bloch's theorem

The periodicity of the solutions of the 1D ring is a simple example of a very general property of eigenstates in an infinite periodic potential.

The Hamiltonian for a particle in a periodic potential

$$H\psi(\mathbf{r}) = \left[-\hbar^2 \nabla^2 / 2m + U(\vec{r})\right]\psi(\mathbf{r}) = E\psi(\mathbf{r}), \qquad (5.13)$$

where $U(\mathbf{r} + \mathbf{R}) = U(\mathbf{r})$ for all \mathbf{R} in a Bravais lattice.

Bloch's theorem states that they have the form

$$\psi_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n\mathbf{k}}(\mathbf{r}) \tag{5.14}$$

where

$$u_{n\mathbf{k}}(\mathbf{r} + \mathbf{R}) = u_{n\mathbf{k}}(\mathbf{r}) \tag{5.15}$$

or, alternatively, that

$$\psi_{n\mathbf{k}}(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}}\psi_{n\mathbf{k}}(\mathbf{r}) \tag{5.16}$$

Notice that while the potential is periodic, the wave function consists of a plane wave times a periodic function. n is an index, call the band index, and we shall see the physical meaning of both n and \mathbf{k} in a moment.

Proof of Bloch's theorem

Here we sketch a proof of Bloch's theorem, and we shall give a somewhat more intuitive (but longer) one later.

First, let us define a *translation operator* $T_{\mathbf{R}}$, which when operating on any function, shifts the argument by a lattice vector \mathbf{R} :

$$T_{\mathbf{R}}f(\mathbf{r}) = f(\mathbf{r} + \mathbf{R}) \tag{5.17}$$

It is straightforward³ to then show that $T_{\mathbf{R}}$ commutes with the Hamiltonian:

$$T_{\mathbf{R}}H = HT_{\mathbf{R}} \tag{5.18}$$

Furthermore

$$T_{\mathbf{R}}T_{\mathbf{R}'} = T_{\mathbf{R}'}T_{\mathbf{R}} = T_{\mathbf{R}+\mathbf{R}'} \quad ; \tag{5.19}$$

the translation operators commute with themselves.

We may now use a fundamental theorem of quantum mechanics; two commuting operators can be chosen to have the same eigenstates, so

$$H\psi = E\psi$$

$$T_{\mathbf{R}}\psi = c(\mathbf{R})\psi$$
(5.20)

Applying the results of Eq. (5.18) and Eq. (5.19), we see that the eigenvalues of T must satisfy

$$c(\mathbf{R})c(\mathbf{R}') = c(\mathbf{R} + \mathbf{R}') \tag{5.21}$$

Now let \mathbf{a}_i be three primitive vectors of the lattice, and write

$$c(\mathbf{a}_i) = e^{2\pi i x_i} \tag{5.22}$$

which is just a definition of the x_i , but we have chosen this form because the boundary conditions will in the end force the x_i to be real. Since a general Bravais lattice vector can be written as $\mathbf{R} = n_1 \mathbf{a_1} + n_2 \mathbf{a_2} + n_3 \mathbf{a_3}$, we can then use the rule of Eq. (5.21) to show that

$$c(\mathbf{R}) = c(\mathbf{a_1})^{n_1} c(\mathbf{a_2})^{n_2} c(\mathbf{a_3})^{n_3} = e^{2\pi i (x_1 n_1 + x_2 n_2 + x_3 n_3)}$$
(5.23)

which is precisely of the form $c(\mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}}$ when

$$\mathbf{k} = x_1 \mathbf{b}_1 + x_2 \mathbf{b}_2 + x_3 \mathbf{b}_3 \tag{5.24}$$

and the \mathbf{b}_i are reciprocal lattice vectors that satisfy $\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi \delta_{ij}$.

This is precisely Bloch's theorem in the form given in Eq. (5.14).

Qu.5.12 * Another proof of Bloch's theorem

A more elegant way to prove Bloch's theorem is to note that the translation operator can be written

$$T_{\mathbf{R}} = e^{-i\mathbf{\hat{P}}\cdot\mathbf{R}/\hbar}$$

where $\hat{\mathbf{P}}$ is the momentum operator. (If you don't know how to do this, make the replacement $\hat{\mathbf{P}} = -i\hbar\nabla$, and check that the operator generates the infinite Taylor series expansion of $f(\mathbf{r} + \mathbf{R})$.) By multiplying by the ket $< \mathbf{k}|$ (an eigenfunction of momentum), show that either $< \mathbf{k}|\psi >= 0$, or $c(\mathbf{R}) = e^{-i\mathbf{k}\cdot\mathbf{R}}$.

 $^{^3 \}textsc{Operate}$ with the translation operator on $H\psi$ and use the periodic symmetry of the potential

5.5 Linear chain revisited

We can use Bloch's theorem to revisit the linear chain problem. If we want to make up a wave-function using only one-orbital per unit cell we now know that it *must* be of the form

$$\psi_k(r) = \frac{1}{\sqrt{N}} \sum_j e^{ikR_j} \phi(r - R_j)$$
(5.25)

and by comparison to the result of explicit calculation in Eq. 5.11 we can connect the notation:

$$R_j = ja; \quad k = 2\pi \frac{m}{Na}; \quad a = \text{lattice constant}$$
 (5.26)

Translational symmetry restricts the phase relationship from site to site. With one orbital per unit cell there is then no choice in the wavefunction. Now evaluate the energy

$$E(k) = \langle \psi_k | H | \psi_k \rangle \tag{5.27}$$

$$= \frac{1}{N} \sum_{j,m} e^{-ikR_j} \langle \phi_j | H | \phi_m \rangle e^{ikR_m}$$
(5.28)

and then writing $R_m = R_j + R_n$

$$E(k) = \frac{1}{N} \sum_{j,n} e^{ikR_n} \times \langle \phi_j | H | \phi_{j+n} \rangle$$
(5.29)

$$= \frac{1}{N} \sum_{j} \sum_{n} e^{ikR_n} \times \langle \phi_0 | H | \phi_n \rangle$$
 (5.30)

$$= t(e^{ika} + e^{-ika}) = 2t\cos ka$$
 (5.31)

k has to be restricted to a window to avoid double counting. The values of k are discrete, but very close together, spaced by $\Delta k = 2\pi/L$, where L = Na. The range of k must cover $k_{max} - k_{min} = 2\pi/a$ to give N states. Frequently, its convenient to choose the range $-\pi/a < k < \pi/a$, the first Brillouin zone.

5.6 LCAO method in general

It is now fairly clear how to extend this method to higher dimensions, and to multiple orbitals per atom, and to multiple unit cells, with wavefunctions of the form

$$\psi_{m,\mathbf{k}}(\mathbf{r}) = \frac{1}{N^{\frac{1}{2}}} \sum_{n=1}^{n_{orb}} \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} c_n^{(m)}(\mathbf{k}) \phi_n(\mathbf{r}-\mathbf{R}) .$$
(5.32)

• **R** are the lattice vectors. ϕ_n is the n^{th} orbital in each unit cell.

- m is the band index: if we have n_{orb} basis functions, we will have $m = 1, ..., n_{orb}$ bands.
- **k** is a three-dimensional vector, now restricted to a three-dimensional Brillouin zone (the Wigner-Seitz cell of the reciprocal lattice)
- $c_n^{(m)}(\mathbf{k})$ is then the matrix of coefficients to be determined by diagonalising the Hamiltonian. The dimension of the matrix concerned is $n_{orb} \times n_{orb}$.

In practice, how to do this:

• Write down a Bloch state made up out of a single orbital in each unit cell.

$$\phi_{n,\mathbf{k}}(\mathbf{r}) = \frac{1}{N^{\frac{1}{2}}} \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \phi_n(\mathbf{r}-\mathbf{R}) .$$
 (5.33)

• Bloch states of different k are orthogonal

$$\left\langle \phi_{m,\mathbf{k}'} | \phi_{n,\mathbf{k}} \right\rangle = \frac{1}{N} \sum_{\mathbf{R},\mathbf{R}'} e^{i\mathbf{k}\cdot\mathbf{R}-i\mathbf{k}'\cdot\mathbf{R}'} \left\langle \phi_m(\mathbf{r}-\mathbf{R}') | \phi_n(\mathbf{r}-\mathbf{R}) \right\rangle$$
(5.34)

$$= \frac{1}{N} \sum_{\mathbf{R},\mathbf{R}''} e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}} e^{i\mathbf{k}'\cdot\mathbf{R}''} \left\langle \phi_m(r-\mathbf{R}-\mathbf{R}'') | \phi_n(\mathbf{r}-\mathbf{R}) \right\rangle 5.35)$$

$$= \delta(\mathbf{k} - \mathbf{k}') \sum_{\mathbf{R}''} e^{i\mathbf{k}' \cdot \mathbf{R}''} < \phi_m(\mathbf{r} - \mathbf{R}'') |\phi_n(\mathbf{r}) >$$
(5.36)

$$\approx \quad \delta(\mathbf{k} - \mathbf{k}')\delta_{m,n} \sum_{\mathbf{R}''} e^{i\mathbf{k}' \cdot \mathbf{R}''} \delta_{\mathbf{R}'',0} = \delta(\mathbf{k} - \mathbf{k}')\delta_{m,n}$$
(5.37)

where the very last line can, in principle, be made exact by choosing a basis of states that are orthogonal in real space (and then known as *Wannier functions*).

• This forms a basis to solve the problem, i.e. the eigenstates are now known to be linear combinations

$$\psi_{m,\mathbf{k}}(\mathbf{r}) = \sum_{n} c_n^{(m)}(\mathbf{k})\phi_{n,\mathbf{k}}(\mathbf{r})$$
(5.38)

- The eigenstates and eigenvalues are determined by the diagonalisation of the matrix $H_{m,n}({\bf k}),$ which has elements

$$H_{m,n}(\mathbf{k}) = \left\langle \phi_{m,\mathbf{k}} \left| H \right| \phi_{n,\mathbf{k}} \right\rangle$$
(5.39)

$$= \frac{1}{N} \sum_{\mathbf{R},\mathbf{R}'} e^{-i\mathbf{k}\cdot\mathbf{R}'} \left\langle \phi_m(\mathbf{r}-\mathbf{R}') | H | \phi_n(\mathbf{r}-\mathbf{R}) \right\rangle e^{i\mathbf{k}\cdot\mathbf{R}}$$
(5.40)

$$= \frac{1}{N} \sum_{\mathbf{R}} \sum_{\mathbf{R}''} e^{-i\mathbf{k}\cdot\mathbf{R}''} \left\langle \phi_m(\mathbf{r} - \mathbf{R} - \mathbf{R}'') | H | \phi_n(\mathbf{r} - \mathbf{R}) \right\rangle \quad (5.41)$$

$$= \sum_{\mathbf{R}''} e^{-i\mathbf{k}\cdot\mathbf{R}''} \left\langle \phi_m(\mathbf{r}-\mathbf{R}'') | H | \phi_n(\mathbf{r}) \right\rangle$$
(5.42)

- Notice that the phase factor involves $\mathbf{k}\cdot\mathbf{R},$ where \mathbf{R} are lattice vectors, not distances between atoms.

Qu.5.13 Tight binding for BCC and FCC lattices

Show that the tightbinding bandstructure based on a single orbital per site for a body centred cubic lattice (include only the hopping to the eight nearest neighbours) is

$$E(\mathbf{k}) = \epsilon_0 + 8t\cos(\frac{1}{2}k_x a)\cos(\frac{1}{2}k_y a)\cos(\frac{1}{2}k_z a) , \qquad (5.43)$$

and for the face centred cubic lattice (twelve nearest neighbours)

$$E(\mathbf{k}) = \epsilon_0 + 4t \left[\cos(\frac{1}{2}k_x a)\cos(\frac{1}{2}k_y a) + \cos(\frac{1}{2}k_y a)\cos(\frac{1}{2}k_z a) + \cos(\frac{1}{2}k_z a)\cos(\frac{1}{2}k_z a)\right]$$
(5.44)

5.7 Periodic boundary conditions and counting states in 3 dimensions

We saw that the spacing between k-points in 1D was $2\pi/L$, where L is the linear dimension of the crystal.

• This generalises to 3 dimensions: the volume associated with each k is

$$\Delta \mathbf{k} = \frac{(2\pi)^3}{V} \tag{5.45}$$

with V the volume of the crystal.

- Within each *primitive unit cell* of the reciprocal lattice there are now precisely N allowed values of \mathbf{k} , (N being the number of unit cells in the crystal).
- In practice N is so big that the bands are continuous functions of k and we only need to remember density of states to count.
- The bandstructure is periodic in the reciprocal lattice $E_n(\mathbf{k} + \mathbf{G}) = E_n(\mathbf{k})$ for any reciprocal lattice vector \mathbf{G} . It is sometimes useful to plot the bands in *repeated zones*, but remember that these states are just being relabelled and are not physically different.
- Since electrons are fermions, each k-point can now be occupied by two electrons (double degeneracy for spin). So if we have a system which contains one electron per unit cell (e.g. a lattice of hydrogen atoms), half the states will be filled in the first Brillouin zone. Two electrons per unit cell fills a Brillouin zone's worth of k states.

Periodic boundary conditions and volume per k-point

A formal proof of the number of allowed k-points uses Bloch's theorem, and follows from the imposition of periodic boundary conditions:

$$\psi(\mathbf{r} + N_i \mathbf{a}_i) = \psi(\mathbf{r}) \tag{5.46}$$

where N_i are integers, with the number of primitive unit cells in the crystal being $N = N_1 N_2 N_3$, and \mathbf{a}_i primitive lattice vectors. Applying Bloch's theorem, we have immediately that

$$e^{iN_i\mathbf{k}\cdot\mathbf{a}_i} = 1, \tag{5.47}$$

so that the general form for the allowed Bloch wavevectors is

$$\mathbf{k} = \sum_{i}^{3} \frac{m_i}{N_i} \mathbf{b}_i, \qquad \text{for } m_i \text{ integral.}$$
(5.48)

with \mathbf{b}_i primitive reciprocal lattice vectors. Thus the volume of allowed k-space per allowed k-point is just

$$\Delta k = \frac{\mathbf{b}_1}{N_1} \cdot \frac{\mathbf{b}_2}{N_2} \wedge \frac{\mathbf{b}_3}{N_3} = \frac{1}{N} \mathbf{b}_1 \cdot \mathbf{b}_2 \wedge \mathbf{b}_3.$$
(5.49)

Since $\mathbf{b}_1 \cdot \mathbf{b}_2 \wedge \mathbf{b}_3 = (2\pi)^3 N/V$ is the volume of the unit cell of the reciprocal lattice (V is the volume of the crystal), Eq. (5.49) shows that the number of allowed wavevectors in the primitive unit cell is equal to the number of lattice sites in the crystal. We may thus rewrite Eq. (5.49)

$$\Delta k = \frac{(2\pi)^3}{V} \tag{5.50}$$

52

Chapter 6

Electronic structure from plane waves

6.1 Nearly free electrons

The tight binding method is clearly the appropriate starting point for a theory when the atomic potential is very strong, and the hopping probability for an electron to move from site-to-site is very small. Here we explore the other limit, where instead the lattice potential is assumed to be weak, and the kinetic energy is the most important term.

6.2 The Schrödinger equation in momentum space

We need to solve

$$H\psi(\mathbf{r}) = \left[-\hbar^2 \nabla^2 / 2m + U(\mathbf{r})\right]\psi(\mathbf{r}) = E\psi(\mathbf{r})$$
(6.1)

We expand the wavefunction in terms of a set of plane waves that satisfy the periodic boundary conditions. 1

$$\psi(\mathbf{r}) = \sum_{\mathbf{k}} c_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}},\tag{6.2}$$

The periodic potential $U(\mathbf{r})$ has a plane wave expansion that only contains waves with the periodicity of the reciprocal lattice

$$U(\mathbf{r}) = \sum_{\mathbf{G}} U_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}},\tag{6.3}$$

¹The sum here is over all k, not just the Brillouin zone. Also the k-points are so dense that we can usually replace the sum by an integral, i.e. $\sum_k \to V/(2\pi)^3 \int d\mathbf{k}$

where \mathbf{G} are the reciprocal lattice vectors.² The momentum components are

$$U_{\mathbf{G}} = \frac{N}{V} \int_{unit\,cell} d\mathbf{r} \, e^{-i\mathbf{G}\cdot\mathbf{r}} \, U(\mathbf{r}) \,, \tag{6.4}$$

and since the potential is real $U_{\mathbf{G}}^* = U_{-\mathbf{G}}$.

We now insert Eq. (6.2) and Eq. (6.3) in Eq. (6.1), and obtain, after a little reorganisation of the terms

$$\sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \left[\left(\frac{\hbar^2}{2m} k^2 - E \right) c_{\mathbf{k}} + \sum_{\mathbf{G}} U_{\mathbf{G}} c_{\mathbf{k}-\mathbf{G}} \right] = 0$$
(6.5)

Since the plane waves form an orthogonal set, each coefficient in the sum over \mathbf{k} in Eq. (6.5) must vanish, i.e.

$$\left[\left(\frac{\hbar^2}{2m}k^2 - E\right)c_{\mathbf{k}} + \sum_{\mathbf{G}}U_{\mathbf{G}}c_{\mathbf{k}-\mathbf{G}}\right] = 0$$
(6.6)

It is often convenient to rewrite $\mathbf{q} = \mathbf{k} - \mathbf{K}$, where **K** is a reciprocal lattice vector chosen so that **q** lies in the first Brillouin zone, when Eq. (6.6) is just

$$\left[\left(\frac{\hbar^2}{2m} (\mathbf{q} - \mathbf{K})^2 - E \right) c_{\mathbf{q} - \mathbf{K}} + \sum_{\mathbf{G}} U_{\mathbf{G} - \mathbf{K}} c_{\mathbf{q} - \mathbf{G}} \right] = 0$$
(6.7)

The wavefunction is of the Bloch form, because Eq. (6.7) mixes plane waves of momentum \mathbf{q} with $\mathbf{q} - \mathbf{G}$, and so

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{k}-\mathbf{G}} e^{i(\mathbf{k}-\mathbf{G})\cdot\mathbf{r}} = e^{i\mathbf{k}\cdot\mathbf{r}} u(\mathbf{r}) , \qquad (6.8)$$

where

$$u(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{k}-\mathbf{G}} e^{-i\mathbf{G}\cdot\mathbf{r}}$$
(6.9)

is now a function with the periodicity of the lattice.

6.3 One-dimensional chain

Let's get some insight from a simple model of a one-dimensional chain, but now simplifying the atomic potential so it just contains the leading Fourier components, i.e

$$U(r) = \frac{1}{2}U_0 \cos\frac{2\pi x}{a}$$
(6.10)

54

 $^{^{2}}$ In contrast to the sum over k, this sum really is discrete.



Figure 6.1: Band structure in the "empty lattice", where U = 0, but we pretend to keep Bloch's theorem, so we have multiple bands in each Brillouin zone. Turning on the potential U_0 splits the degeneracies and opens up band gaps

Then the secular determinant runs just down the tri-diagonal³

The physical interpretation of this is that an incident plane wave with wavevector k can be scattered by the potential into a state of $k \pm 2\pi/a$. Multiple scattering then mixes these terms together. If U_0 is small, we should be able to treat it perturbatively, remembering to take care of degeneracies. Of course if $U_0 = 0$, we get the free electron eigenvalues

$$E_0^{(m)}(k) = (k - 2\pi m/a)^2 \quad m = \dots, -2, -1, 0, 1, 2, \dots$$
(6.12)

which are now repeated, offset parabolas. Remember Bloch's theorem tells us that k as a crystal momentum is conserved only within the first Brillouin zone (see Fig. 6.1).

Now suppose U_0 is turned on, but is very small. It will be important only for those momenta when two free electron states are nearly degenerate, for example, m=0,1 are degenerate when $k = \pi/a$. Near that point, we can simplify the band

³We set $\hbar^2/2m = 1$ for a moment



Figure 6.2: Energy bands in one dimension, and amplitudes of wavefunctions at the band edges

structure to the 2x2 matrix

$$\begin{pmatrix} (k - \frac{2\pi}{a})^2 - E & U_0 \\ U_0 & k^2 - E \end{pmatrix} \begin{pmatrix} c_1 \\ c_0 \end{pmatrix}$$
(6.13)

The solution of the determinantal leads to a quadratic equation:

$$E^{\pm}(\mathbf{k}) = \frac{1}{2}(k^2 + (k - 2\pi/a)^2) \pm \frac{1}{2}\sqrt{(k^2 - (k - 2\pi/a)^2)^2 + 4U_0^2}$$
(6.14)

Exactly at $k = \pi/a$, the energy levels are

$$E^{\pm}(\pi/a) = E^{0}_{\pi/a} \pm |U_0|, \qquad (6.15)$$

and if we choose the potential to be attractive $U_0 < 0$, the wavefunctions are (aside from normalisation)

$$\psi^{-}(\pi/a) = \cos(\pi x/a),
\psi^{+}(\pi/a) = \sin(\pi x/a).$$
(6.16)

The wavefunctions are plotted, along with the potential, in Fig. 6.2.



Figure 6.3: Diatomic chain of atoms. Note that if the potentials on the two atoms are identical, and $\delta = 0$, the chain converts to a monatomic chain of period a/2

Qu.6.14 The diatomic chain Consider a diatomic lattice of two atoms labelled A and B in a lattice with period a, at the positions $\pm a/4(1-\delta)$ in a one-dimensional array with overall period a (see Fig. 6.3).

Using the NFE approximation valid for momenta near the zone boundary $k \to \pi/a$, show that the solution of Eq. (6.14) leads to: (a) a gap on the zone boundary of magnitude $2|U_{2\pi/a}|$, and (b) wavefunctions that satisfy $c_k^{\pm}/c_{k-2\pi/a}^{\pm} = \pm U/|U|$ as $\mathbf{k} \to \pi/a$.

Hence show that the probability density for the electronic states at $k=\pi/a$ take the form

$$|\psi^{+}(r)|^{2} \propto \cos^{2}(\frac{\pi x}{a} + \frac{\phi}{2})$$

 $|\psi^{-}(r)|^{2} \propto \sin^{2}(\frac{\pi x}{a} + \frac{\phi}{2}).$ (6.17)

Show that the potential can be written

$$U_{2\pi/a} = \sin(\frac{\pi\delta}{2})(U_{2\pi/a}^A + U_{2\pi/a}^B) - i\cos(\frac{\pi\delta}{2})(U_{2\pi/a}^A - U_{2\pi/a}^B), \qquad (6.18)$$

where

$$U_{2\pi/a}^{A,B} = \frac{1}{L} \int dr \, e^{-2\pi i r/a} U^{A,B}(r) \,, \tag{6.19}$$

(here $L \to \infty$ is the length of the chain).

The system contains an average of one electron per atom, or equivalently two electrons per unit cell. Discuss the values of the energy gaps and plot the charge densities corresponding to the highest filled electron state and the lowest empty electron state in the three cases; (a) identical atoms, $U_A = U_B$, and $\delta = 0$;

- (b) different atoms $U_A \neq U_B$, and $\delta = 0$, ;
- (c) identical atoms, $U_A = U_B$, and $\delta \neq 0$.

Explain how this provides a simple model of either an *ionic* or *covalent* solid.

6.4 Pseudopotential

The NFE method and the tight-binding method are not accurate methods of electronic structure determination; nevertheless both of them exhibit the basic principles. They are commonly used to write down simple models for bands, with their parameters fit to more sophisticated calculations, or to experiment. It turns out that band gaps in semiconductors are usually fairly small, and the true dispersion can be modelled by scattering from a few Fourier components of the lattice potential. The reason is that the relevant scattering potential for valence band electrons is however MUCH smaller than the full atomic potential ze^2/r of an electron interacting with a nucleus of charge z. The effective potential for scattering of the valence electrons by the atomic cores is a weak *pseudopotential*.

When we consider the band structure of a typical solid, we are concerned only with the valence electrons, and not with those tightly bound in the core, which remain nearly atomic. If we solve the full Schrödinger equation with the real Coulomb potential, we expect to calculate not just the valence electronic states, but also the atomic like core states. A pseudopotential reproduces the valence states as the *lowest* eigenstates of the problem and neglects the core states.



Figure 6.4: Pseudopotential: The true potential V(r) has a wavefunction for the valence electrons that oscillates rapidly near the core. The pseudopotential $V_s(r)$ has a wavefunction $\Phi_s(r)$ that is smooth near the core, but approximates the true wavefunction far from the core region.

A weak pseudopotential acting on a smooth pseudo-wavefunction gives nearly

the same energy eigenvalues for the valence electrons as the full atomic potential does acting on real wavefunctions. Away from the atomic cores, the pseudopotential matches the true potential, and the pseudo-wavefunction approximates the true one.

A formal derivation of how this works can be given using the method of orthogonalised plane waves. The atomic states are well described by the Bloch functions $f_{n\mathbf{k}}$ of the LCAO or tight-binding scheme Eq. (5.32). Higher states, which extend well beyond the atoms will not necessarily be of this kind, but they must be *orthogonal* to the core levels. This suggests that we should use as a basis ⁴

$$|\chi_{\mathbf{k}}\rangle = |\mathbf{k}\rangle - \sum_{n} \beta_{n} |f_{n\mathbf{k}}\rangle, \qquad (6.20)$$

where $|\mathbf{k}\rangle$ is a plane wave, and the coefficients $\beta_n(\mathbf{k})$ are chosen to make the states χ orthogonal to the core states $|f_{n\mathbf{k}}\rangle$. The states in Eq. (6.20) are *orthogonalised plane waves* (OPW); away from the core, they are plane wave like, but in the vicinity of the core they oscillate rapidly so as to be orthogonal to the core levels.

We can now use the OPW's as basis states for the diagonalisation in the same way that we used plane waves in the NFE, viz

$$|\psi_k\rangle = \sum_{\mathbf{G}} \alpha_{\mathbf{k}-\mathbf{G}} |\chi_{\mathbf{k}-\mathbf{G}}\rangle .$$
(6.21)

This turns out to converge very rapidly, with very few coefficients, and only a few reciprocal lattice vectors are included in the sum. The following discussion explains why.

Suppose we have solved our problem exactly and determined the coefficients α . Now consider the sum of plane waves familiar from the plane-wave expansion, but using the same coefficients, i.e.

$$|\phi_k\rangle = \sum_{\mathbf{G}} \alpha_{\mathbf{k}-\mathbf{G}} |\mathbf{k}-\mathbf{G}\rangle, \qquad (6.22)$$

and then 5 it is easily shown that

$$|\psi> = |\phi> -\sum_{n} < f_{n}|\phi>|f_{n}>$$
 (6.23)

Then substitute into the Schrödinger equation $H|\psi\rangle = E|\psi\rangle$, which gives us

$$H|\phi> + \sum_{n} (E - E_n) < f_n|\phi> |f_n> = E|\phi>$$
 (6.24)

We may look upon this as a new Schrödinger equation with a pseudopotential defined by the operator

$$V_s |\phi\rangle = U |\phi\rangle + \sum_n (E - E_n) < f_n |\phi\rangle |f_n\rangle$$
 (6.25)

which may be written as a non-local operator in space

$$(V_s - U)\phi(r) = \int V_R(\mathbf{r}, \mathbf{r}')\phi(\mathbf{r}') \, d\mathbf{r}' \quad , \tag{6.26}$$

⁴We use Dirac's notation of *bra* and *ket*, where $|\mathbf{k}\rangle$ represents the plane wave state $\exp(i\mathbf{k}\cdot\mathbf{r})$, and $\langle \phi_1|T|\phi_2\rangle$ represents the matrix element $\int d\mathbf{r} \phi_1^*(\mathbf{r})T(\mathbf{r})\phi_2(\mathbf{r})$ of the operator T.

 $^{^5\}mathrm{Saving}$ more notation by dropping the index k

where

$$V_R(\mathbf{r}, \mathbf{r}') = \sum_n (E - E_n) f_n(\mathbf{r}) f_n^*(\mathbf{r}') . \qquad (6.27)$$

The pseudopotential acts on the smooth *pseudo-wavefunctions* $|\phi\rangle$, whereas the bare Hamiltonian acts on the highly oscillating wavefunctions $|\psi\rangle$.

One can see in Eq. (6.25) that there is cancellation between the two terms. The bare potential is large and attractive, especially near the atomic core at $r \approx 0$; the second term V_R is positive, and this cancellation reduces the total value of V_s especially near the core. Away from the core, the pseudopotential approaches the bare potential.

Qu.6.15 Nearly free electron approximation for a square lattice The potential in a 2-dimensional square crystal of side a is given by

$$V(x,y) = -2V_0 \left[\cos\left(\frac{2\pi x}{a}\right) + \cos\left(\frac{2\pi y}{a}\right) \right] \quad . \tag{6.28}$$

Use the nearly-free electron approximation to calculate the electron energies at the wave-vectors

$$\mathbf{k}_0 = \frac{2\pi}{a}(0,0) \ , \mathbf{k}_1 = \frac{2\pi}{a}(\frac{1}{2},0) \ , \mathbf{k}_2 = \frac{2\pi}{a}(\frac{1}{2},\frac{1}{2}) \ . \tag{6.29}$$

(a) Write down the form of the wavefunction within the nearly-free-electron approximation, using 1 plane wave at \mathbf{k}_0 , 2 plane waves at \mathbf{k}_1 , and 4 plane waves at \mathbf{k}_2 .

(b) In each case, substitute these wavefunctions into the Schrödinger equation, and write the resulting equations in matrix form.

(c) Solve the three eigenvalue problems for the energy levels at ${\bf k}_0,\, {\bf k}_1,\, \text{and}\,\, {\bf k}_2.$

Chapter 7

Bandstructure of real materials

7.1 Bands and Brillouin zones

In the last chapter, we noticed that we get band gaps forming by interference of degenerate forward- and backward going plane waves, which then mix to make standing waves.

Brillouin zones. What is the condition that we get a gap in a threedimensional band structure? A gap will arise from the splitting of a degeneracy due to scattering from some fourier component of the lattice potential, i.e. that

$$E_0(\mathbf{k}) = E_0(\mathbf{k} - \mathbf{G}) \tag{7.1}$$

which means (for a given **G**) to find the value of k such that $|\mathbf{k}|^2 = |\mathbf{k} - \mathbf{G}|^2$. Equivalently, this is

$$\mathbf{k} \cdot \frac{\mathbf{G}}{2} = \left| \frac{\mathbf{G}}{2} \right|^2 \tag{7.2}$$

which is satisfied by any vector lying in a plane perpendicular to, and bisecting **G**. This is, by definition, the boundary of a Brillouin zone; it is also the Bragg scattering condition, not at all coincidentally.

Electronic bands. We found that the energy eigenstates formed discrete bands $E_n(\mathbf{k})$, which are *continuous* functions of the momentum \mathbf{k} and are additionally labelled by a *band index n*. The bands are periodic: $E_n(\mathbf{k}+\mathbf{G}) = E_n(\mathbf{k})$.

Bloch's theorem again. The eigenstates are of the form given by Bloch's theorem

$$\psi_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n\mathbf{k}}(\mathbf{r}) \tag{7.3}$$

where $u(\mathbf{r})$ is periodic on the lattice. Notice that if we make the substitution $\mathbf{k} \to \mathbf{k} + \mathbf{G}$, Eq. (7.3) continues to hold. This tells us that \mathbf{k} can always be

chosen inside the first Brillouin zone for convenience, although it is occasionally useful to plot the bands in an extended or repeated zone scheme as in Fig. 6.2.

Crystal momentum. The quantity $\hbar \mathbf{k}$ is the crystal momentum, and enters conservation laws for scattering processes. For example, if an electron absorbs the momentum of a phonon of wavevector \mathbf{q} , the final state will have a Bloch wavevector $\mathbf{k}' = \mathbf{k} + \mathbf{q} + \mathbf{G}$, where \mathbf{G} is whatever reciprocal lattice vector necessary to keep \mathbf{k}' inside the Brillouin zone. Physical momentum can always be transferred to the lattice in arbitrary units of $\hbar \mathbf{G}$. Notice that depending on the energy conservation, processes can thus lead to transitions between bands.

Counting states. In a big system, the allowed k-points are discrete but very closely spaced. Each occupies a volume

$$\Delta \mathbf{k} = \frac{(2\pi)^3}{V} \tag{7.4}$$

with V the volume of the crystal. Thus within each *primitive unit cell* or *Brillouin zone* of the reciprocal lattice there are now precisely N allowed values of \mathbf{k} , (N being the number of unit cells in the crystal).

Even number rule. Allowing for spin, two electrons per real space unit cell fills a Brillouin zone's worth of k states.

7.2 Metals and insulators in band theory

The last point is critical to the distinction that band theory makes between a metal and an insulator. A (non-magnetic) system with an even number of electrons per unit cell *may* be an insulator. In all other cases, the fermi energy must lie in a band and the material will be predicted¹ to be a metal. Metallicity may also be the case even if the two-electron rule holds, if different bands overlap in energy so that the counting is satified by two or more partially filled bands.

Notation

The bandstructure $E_n(\mathbf{k})$ defines a function in three-dimensions which is difficult to visualise. Conventionally, what is plotted are cuts through this function along particular directions in k-space. Also, a shorthand is used for directions in k-space and points on the zone boundary, which you will often see in band structures.

- $\Gamma = (0, 0, 0)$ is the zone centre.
- X is the point on the zone boundary in the (100) direction; Y in the (010) direction; Z in the (001) direction. *Except* if these directions are equivalent by symmetry (e.g. cubic) they are all called "X".

¹Band theory may fail in the case of *strongly correlated systems* where the Coulomb repulsion between electrons is larger than the bandwidth, producing a *Mott insulator*

- L is the zone boundary point in the (111) direction.
- K in the (110) direction.
- You will also often see particular bands labelled either along lines or at points by greek or latin capital letters with a subscript. These notations label the group representation of the state (symmetry) and we won't discuss them further here.

Density of states

We have dealt earlier with the density of states of a free electron band in 3, especially Question 3.2. The maxima E_{max} and minima E_{min} of all bands must have a locally quadratic dispersion with respect to momenta measured from the minima or maxima. Hence the density of states (in 3D) near the minima will be the same

$$g(E \stackrel{>}{\sim} E_{min}) = \frac{V}{\pi^2} \frac{m^*}{\hbar^2} \left(\frac{2m^*(E - E_{min})}{\hbar^2}\right)^{\frac{1}{2}} \quad . \tag{7.5}$$

as before, with now however the replacement of the bare mass by an effective mass $m^* = (m_x^* m_y^* m_z^*)^{1/3}$ averaging the curvature of the bands in the three directions. A similar form must apply near the band maxima, but with now $g(E) \propto (E_{max} - E)^{\frac{1}{2}}$. Notice that the flatter the band, the larger the effective mass, and the larger the density of states².

Since every band is a surface it will have saddle points (in two dimensions or greater) which are points where the bands are flat but the curvature is of opposite signs in different directions. Examples of the generic behaviour of the density of states in one, two and three dimensions are shown in Fig. 7.1. The saddle points give rise to cusps in the density of states in 3D, and a logarithmic singularity in 2D.

For any form of E(k), the density of states is

$$g(E) = \sum_{n} g_n(E) = \sum_{n} \int \frac{d\mathbf{k}}{4\pi^3} \delta(E - E_n(\mathbf{k})) , \qquad (7.6)$$

Because of the δ -function in Eq. (7.6), the momentum integral is actually over a surface in k-space S_n which depends on the energy E; $S_n(E_F)$ is the Fermi surface. We can separate the integral in \mathbf{k} into a two-dimensional surface integral along a contour of constant energy, and an integral perpendicular to this surface dk_{\perp} (see Fig. 7.2). Thus

$$g_n(E) = \int_{S_n(E)} \frac{dS}{4\pi^3} \int dk_{\perp}(\mathbf{k}) \,\delta(E - E_n(\mathbf{k}))$$
$$= \int_{S_n(E)} \frac{dS}{4\pi^3} \frac{1}{|\nabla_{\perp} E_n(\mathbf{k})|},$$
(7.7)

 $^{^{2}}$ The functional forms are different in one and two dimensions, see Question 3.2.



Figure 7.1: Density of states in one (top curve), two (middle curve) and three (lower curve) dimensions



Figure 7.2: Surface of constant energy

where $\nabla_{\perp} E_n(\mathbf{k})$ is the derivative of the energy in the normal direction.³

Notice the appearance of the gradient term in the denominator of Eq. (7.7), which must vanish at the edges of the band, and also at saddle points, which exist generically in two and three dimensional bands. Maxima, minima, and saddle points are all generically described by dispersion (measured relative to the critical point) of

$$E(\mathbf{k}) = E_0 \pm \frac{\hbar^2}{2m_x} k_x^2 \pm \frac{\hbar^2}{2m_y} k_y^2 \pm \frac{\hbar^2}{2m_z} k_z^2$$
(7.8)

If all the signs in Eq. (7.8) are positive, this is a band minimum; if all negative, this is a band maximum; when the signs are mixed there is a saddle point. In the vicinity of each of these critical points, also called van Hove singularities, the density of states (or its derivative) is singular. In two dimensions, a saddle point gives rise to a logarithmically singular density of states, whereas in three dimensions there is a discontinuity in the derivative.

Qu.7.16 2D tight binding band

This question is to encourage you to visualise bands in two dimensions (and higher!). Using a simple numerical package to plot representative cases will help.

Consider a two-dimensional band structure on a rectangular lattice

$$E(k) = -2 * t_1 \cos(ak_x) - 2 * t_2 \cos(bk_y)$$
(7.9)

(a) What is the reciprocal lattice? Draw the first Brillouin zone boundary.

(b) What is the real space lattice?

(c) Suppose that $t_1 > t_2 > 0$ and a < b. (Do you expect there to be a relation?) Plot some contours of constant energy. At which momenta do you find the band minima, maxima, and saddle points? What are the effective masses of electrons at these points (keep track of signs)?

(d) Make a numerical estimation of the density of states as a function of energy (plot a histogram, say). Can you give an analytic form for the energy- dependence of the density of states near the singular points?

(e) For what range of energies are the energy contours open or closed? Does this bear any relationship to the energies of the saddle points?

³We are making use of the standard relation $\delta(f(x) - f(x_0)) = \delta(x - x_0)/|f'(x_0)|$



Figure 7.3: Band structure of Al (solid line) compared to the free electron parabolas (dotted line). Calculations from Stumpf and Scheffler, cited by Marder.

7.3 Examples of band structures

Metals

If there are not an even number of electrons per unit cell, the chemical potential must lie in a band, and there will be no energy gap. Because there are low-lying electronic excitations, the system is a metal. The *fermi surface* is the surface in momentum space that separates the filled from the empty states. In a simple metal like Na ($3s^1 - 1$ valence electron) or Al ($3s^2p^1 - 3$ valence electrons) this is nearly a sphere like the free electron gas. In other cases (e.g. Cu, $4s3d^{10}$) the sphere extends in some directions to meet the Brillouin zone boundary surface. There can be situations where several bands are cut by the fermi energy, and the topology of fermi surfaces is sometime complicated.

Semimetals

Even if there are the right number of electrons to fill bands and make a semiconductor, the bands may still overlap. Consequently, the chemical potential will intersect more than one band, making a pocket of electrons in one band and removing a pocket of electrons from the band below (which as we shall see later, are sometimes called holes). This accounts for the metallicity of Ca and Mg(which have two electrons per unit cell), and also As, Sb and Bi. The latter, despite being group V elements, have crystal structures that contain 2 atoms per unit cell and therefore 10 valence electrons. We have previously alluded

66



Figure 7.4: Band structure of Cu metal [from G.A.Burdick, *Phys. Rev.***129**,138 (1963)], cited by Grosso and Parravicini



Figure 7.5: Fermi surface of Cu



Figure 7.6: Pseudopotential band structure of Si and Ge [M.L.Cohen and T.K.Bergstresser *Phys.Rev*141, 789 (1966)]. The energies of the optical transitions are taken from experiment.

to graphite, which is a special kind of semimetal. We noted that a graphene sheet has conduction and valence bands that touch at special points on the zone boundary. Over all except these points, the band structure has a gap - thus graphene is more correctly titled a *zero-gap semiconductor*.

Semiconductors and insulators

68

If there are an even number of electrons per unit cell, then it is possible (if the bands don't overlap) for the occupied states all to lie in a set of filled bands, with an energy gap to the empty states. In this case the system will be a semiconductor or insulator. Such is the case for the group IV elements C, Si and Ge, as well as the important III-V compounds such as GaAs and AlAs. These elements and compounds in fact have 2 atoms per unit cell (diamond or zincblende structure) and have a total of 8 valence electrons per unit cell — 4 filled bands.

The band structures of Si, Ge, and GaAs are shown in Fig. 7.6 and Fig. 7.7. The maximum of the valence bands of all the materials is at Γ . Si and Ge are both *indirect* gap materials, because the conduction bands have minima either in the (100) direction (Si) or the (111) direction (Ge).



Figure 7.7: Band structure of GaAs [M.L.Cohen and T.K.Bergstresser *Phys.Rev*141, 789 (1966)]



Figure 7.8: The valence charge density for Ge, GaAs, and ZnSe from an early pseudopotential calculation, plotted along a surface in a 110 plane that contains the two atoms of the unit cell. Note the (pseudo-)charge density shifting from the centre of the bond in Ge to be almost entirely ionic in ZnSe. [M.L.Cohen, *Science* **179**, 1189 (1973)]

7.4 Probing the band structure and density of states

Optical transitions

The band structure provides the excitation spectrum of the solid. The ground state of the system involves filling states up to the fermi energy, but we can also excite the system in different ways. One of the simplest is the absorption of a photon, which can be visualised as an excitation of an electron from an occupied state into an empty state, leaving behind a "hole" in the valence band. See Fig. 7.9.

The minimum gap in a semiconductor is the energy difference between the highest occupied state and the lowest unoccupied state, and this is the threshold for optical absorption (neglecting excitonic physics, see later). In some semiconductors, the maximum valence band state and the minimum in the conduction band occur at the same momentum - in such a *direct gap* system, direct optical excitation is allowed at the minimum gap, and an important example is GaAs.

Si and Ge are example of *indirect gap* materials, because the conduction band minimum is toward the edge of the zone boundary. The minimum energy transition is at large momentum, and therefore cannot be accomplished by direct absorption of a photon. The lowest energy transition is instead a *phonon-mediated* transition where the energy is provided by the photon and the momentum provided by the phonon. This is much less efficient than direct optical absorption.



Figure 7.9: Direct absorption by light is a nearly vertical transition since the wavevector of a photon with energy of order a semiconductor gap is much smaller than the typical momentum of an electron. (a) In a *direct gap* semiconductor, such as GaAs, the lowest energy available states for hole and electron are at the same momentum, and the optical threshold is at the vertical energy gap. (b) IN an indirect gap material (e.g. Si or Ge), the minimum energy excitation of electron and hole pair connects state of different momenta - and a phonon of momentum q must be excited concurrently with the photon.

Luminescence is the inverse process of recombination of an electron-hole



Figure 7.10: The interband absorption spectrum of Si has a threshold at the indirect gap $E_g \approx 1.1 \ eV$ which involves a phonon and is very weak. The energies E_1 and E_2 correspond to critical points where the conduction and valence bands are vertically parallel to one another; absorption is direct (more efficient) and also enhanced by the enhanced joint density of electron and hole states. [E.D.Palik, *Handbook of the optical constants of solids*, AP, 1985].

pair to emit light. It comes about if electrons and holes are injected into a semiconductor (perhaps electrically, as in a light-emitting diode). Obviously, this process will not be efficient in an indirect gap semiconductor but is more so in a direct gap material. This simple fact explains why GaAs and other III-V compounds are the basis of most practical opto-electronics in use today, whereas Si is the workhorse of electrical devices.



Figure 7.11: Schematics of a photoemission experiment. The optical transitions are nearly vertical, so the electrons are excited from the valence bands to high energy excited states (above the vacuum energy necessary to escape from the crystal) with the same crystal momentum. In this case the two transitions that can be excited by a single frequency will yield a double peak in the kinetic energy distribution of the escaped electrons. When the excited electrons escape through the surface of the crystal, their momentum perpendicular to the surface will be changed. If the surface is smooth enough, the momentum of the electron parallel to the surface is conserved, so the angle of the detector can be used to scan k_{\parallel}

Photoemission

The most direct way to measure the electron spectral function directly is by photoemission, although this is a difficult experiment to do with high resolution. In a photoemission experiment, photons are incident on a solid, and cause transitions from occupied states to plane wave-like states well above the vacuum energy; the excited electron leaves the crystal and is collected in a detector that analyses both its energy and momentum.⁴ The photon carries very little momentum, so the momentum of the final electron parallel to the surface is the same as the initial state in the solid, while of course the perpendicular component of the momentum is not conserved. Photoemission data is therefore most easy to interpret when there is little dispersion of the electronic bands perpendicular to the surface, as occurs in anisotropic layered materials. It is fortunate that there are many interesting materials (including high-temperature superoconductors) in this class.

If one analyses both the energy and the momentum of the outgoing electron, (this is Angle Resolved Photo-Emission Spectroscopy, or ARPES) one can de-

72

 $^{^4 {\}rm For}$ a detailed discussion of photoemission experiments, see Z.X.Shen and D.S.Dessau, *Physics Reports*, **253**, 1-162 (1995)


Figure 7.12: Idealised results from a photoemission experiment. A peak is observed at the band energy in each spectrum, but disappears when the band crosses the Fermi energy

termine the band structure directly. Integrating over all angles gives a spectrum that is proportional to the total density of states.

The ideal schematic for interpreting an ARPES experiment would then be as shown in Fig. 7.11. An example of real data is shown in Fig. 7.13.

Photoemission can give information only about occupied states. The technique of *inverse photoemission* involves inserting an electron of known energy into a sample and measuring the ejected photon. Since the added electron must go into unoccupied state, this spectroscopy allows one to map out unoccupied bands, providing information complementary to photoemission.

Tunnelling

Tunnelling spectroscopies (injecting or removing electrons) through a barrier have now evolved to be very important probes of materials. The principle here is that a potential barrier allows one to maintain a probe (usually a simple metal) at an electrical bias different from the chemical potential of the material. Thus the current passed through the barrier comes from a non-equilibrium injection (tunnelling) through the barrier.

A model for a simple metal tunnelling into a more complex material is shown in Fig. 7.14. With the metal and sample maintained at different electrical potentials separated by a bias eV, then the current through the junction can be estimated to be of the form

$$I \propto \int_{\mu+eV}^{\mu} g_L(\omega) g_R(\omega) T(\omega)$$
(7.10)

where T is the transmission through the barrier for an electron of energy ω and



Figure 7.13: . Photoemission spectra on the two dimensional layered metal $SrRuO_4$. The bands are nearly two-dimensional in character, so the interpretation of the photoemission data is straightforward – different angles (see Fig. 7.11) correspond to different in-plane momenta. The upper panels show energy scans for different angles that correspond to changing the in-plane momentum in the direction from the centre of the Brillouin zone Γ towards the centre of the zone face M and the corner X. Several bands cross the Fermi energy, with different velocities, and sharpen as their energies approach E_F . The left hand lower panel plots the positions of the peaks as a function of momentum at the fermi energy, to be compared with the band structure calculation of the fermi surface(s) on the lower right. [Experiment from Damascelli et al, PRL; theory from Mazin et al PRL 79, 733 (1997)]

 g_L and g_R are the densities of states.⁵ If the barrier is very high so that T is not a strong function of energy, and if the density of states in the contact/probe is approximately constant, then the energy-dependence comes entirely from the density of states inside the material. Notice then that the differential conductivity is proportional to the density of states (see Fig. 7.14):

$$dI/dV \propto g(\mu + eV) \quad . \tag{7.11}$$

It is difficult to maintain very large biases, so most experiments are limited to probing electronic structure within a volt or so of the fermi energy.



Figure 7.14: Schematic description of tunnelling between two materials maintained at a relative bias eV. The current is approximately given by the integrated area between the two chemical potentials (provided the matrix element for tunnelling is taken constant.) If the density of states of the contact (or probe, labelled 1 in the figure) is also slowly varying, then the differential conductance dI/dV is proportional to the density of states of the material itself, at the bias eV above the chemical potential μ_2 .

Tunnel junctions are sometimes fabricated by deposition of a thin insulating layer followed by a metal contact.

The technique of scanning tunnelling microscopy (STM) uses a small tip, with vacuum as the surface barrier. Because the tunnel probability is an exponential function of the barrier thickness, this scheme provides high (close to atomic, in some cases) spatial resolution, even though the tip radius will be nmor larger. By hooking this up to a piezoelectric drive in a feedback loop, it has proved possible to provide not only I - V characteristics at a single point, but also spatial maps of the surface.

Scanned probe spectroscopies have advanced to become extraordinary tools at the nanoscale. As well as STM, it is possible to measure forces near a surface

 $^{^5}$ Strictly this formula applies when the tunnelling process does not conserve momentum parallel to the interface, i.e. if the surface is rough or disorded.



Figure 7.15: Differential conductance of a tunnel junction between superconducting Pb and metallic Mg reveals the gap in the density of states of superconducting lead. [I. Giaever, Nobel Prize Lecture, 1973]



Figure 7.16: An array of Fe atoms arranged in a corral on the surface of Cu traps a surface electron state whose density can be imaged by STM. M.F. Crommie, C.P. Lutz, D.M. Eigler, E.J. Heller. Surface Review and Letters 2 (1), 127-137 (1995).

(atomic force microscopy, AFM), which is particularly useful for insulating samples. It has proven possible to manipulate individual atoms, to measure the magnetism of a single spin, and with small single-electron transistors to study to motion of single electron charges in the material.

Qu.7.17 Optical absorption of simple metals

Why are electronic transitions involving photons normally regarded as "vertical" while those involving phonons are "horizontal"?

In the first Brillouin zone of a body centred cubic (BCC) crystal, the shortest distance from the zone centre to the zone boundary is $\sqrt{2}\pi/a$. Show that the free electron fermi surface of a monovalent metal is contained entirely within the first Brillouin zone.

Indicate on a diagram the threshold energy E_0 for the absorption of photons leading to an electron being excited from occupied energy levels to unoccupied energy levels. For a BCC material, show that this energy is $E_0 \approx 0.64 E_F$.

Alkali metals have a BCC structure. The experimental data below show the frequency dependence of the conductivity in the alkali metals Na, K, and Rb, which have lattice constants a, respectively, of 0.423 nm, 0.523 nm, and 0.559 nm. The broad peaks at higher frequencies in each curve have been interpreted as arising from interband optical absorption. Is this consistent with nearly free electron behaviour?



Qu.7.18 * Graphite A single sheet of graphite has two carbon atoms in the unit cell at positions $\mathbf{d}_1 = 0$ and $\mathbf{d}_2 = (a/\sqrt{3})(0, 1, 0)$. The translation vectors for the two-dimensional hexagonal lattice are $\mathbf{t}_1 = (a/2)(1, \sqrt{3}, 0)$ and $\mathbf{t}_1 = (a/2)(-1, \sqrt{3}, 0)$.

The electronic configuration of the carbon atom is $1s^22s^22p^2$, and ignoring the 1s core states, we need to make a band structure from the s, p_x , p_y abd p_z orbitals. Because s, p_x and p_y orbitals are even under reflection through the plane, and p_z odd, the two sets do not mix. The first three states hybridise to form σ -bonds with a large gap between the bonding and anti-bonding orbitals. Within this gap lie the π -orbitals arising from the hybridised p_z . The three bonding σ orbitals will accommodate 6 electrons per cell, leaving 2 electrons per unit cell in the π -bands. This question considers the electronic π -bands only.



Figure 7.17: Two dimensional structure of graphite

(a) Construct Bloch states that consist of a linear mixture of the two p_z orbitals in the unit cell, and show how this gives rise to the secular equation to determine the eigenstate energies

$$\begin{vmatrix} E_p - E & tF(\mathbf{k}) \\ tF^*(\mathbf{k}) & E_p - E \end{vmatrix} = 0 , \qquad (7.12)$$

where $t\ {\rm is}\ {\rm the}\ {\rm two}\ {\rm center}\ {\rm hopping}\ {\rm matrix}\ {\rm element}\ {\rm between}\ {\rm neighbouring}\ p_z\ {\rm orbitals},\ {\rm and}$

$$F(\mathbf{k}) = 1 + 2\cos\left(\frac{k_x a}{2}\right)\exp\left(-i\frac{\sqrt{3}k_y a}{2}\right) .$$
 (7.13)

(b) Show that the reciprocal lattice is also a hexagonal lattice, at an angle of $\pi/6$ to the real-space lattice. Show that the first Brillouin zone is a hexagon centred at the point $\Gamma = (000)$, whose corners are at the points $P = (2\pi/a)(2/3, 0, 0)$

(c) Determine a formula for the dispersion curves for the two eigenstates, and plot them in the directions ΓP , and ΓQ . (Here $Q = (2\pi/a)(1/2, 1/2\sqrt{3}, 0)$ is at the middle of a zone face.

(d) Where will the π -bands lie in energy relative to the $sp^2 \sigma$ -orbitals? Is a single layer of graphite a metal or an insulator?

(e) Carbon nanotubes are formed by curling a graphite sheet into a tube, connecting the atoms with periodic boundary conditions. There are many ways to do this, and the different nanotubes can be indexed by the vector $m\mathbf{t}_1 + n\mathbf{t}_2$ that identifies which atoms are connected periodically. Assuming the band-structure is unchanged, show that the allowed k-states now lie on a set of lines whose direction is parallel to the tube. Discuss the situations under which the resulting tube will be semiconducting or metallic.

Qu.7.19 Band structure of d-band metals

In many transition metals a narrow *d*-band lies within a broad energy band originating from *s*-orbitals. This question discusses the band structure using a simple one-dimensional model contructed from a tight-binding Hamiltonian with one s-orbital $\phi_s(r)$ and one d-orbital $\phi_d(r)$ per atom; the atoms are arranged in a linear chain of lattice constant *a*.

(a) Write down two Bloch states $\phi_s(k)$ and $\phi_d(k)$ formed from the atomic s- and d- states respectively. The eigenstates must be linear combinations of these.

(b) Hence show that the one-particle bandstructure E(k) can be found from the determinantal equation

$$\begin{vmatrix} E_s - 2t_{ss}\cos(ka) - E(k) & -2t_{sd}\cos(ka) \\ -2t_{sd}\cos(ka) & E_d - 2t_{dd}\cos(ka) - E(k) \end{vmatrix} = 0 .$$

Identify and explain the parameters appearing in the determinantal equation, and discuss the approximations made that lead to this form.

(c) Discuss why you would expect that $t_{ss} > |t_{sd}| > t_{dd}$.

(d) Plot the dispersion of the two bands when $|E_d - E_s| \ll 2|t_{ss}|$, and t_{sd} and t_{dd} are neglected.

(e) How is the dispersion modified from (d) by the inclusion of small values of t_{sd} and t_{dd} ?

(f) Discuss the relevance of this model to the electronic bandstructure of Cu metal.

Chapter 8

Electrodynamics of metals

8.1 Screening and Thomas-Fermi theory

One of the most important manifestations of electron-electron interactions in metals is that of screening. If we insert a positive test charge into a metal, it attracts a cloud of electrons around it, so that at large distances away from the test charge the potential is perfectly screened - there is *zero* electric field inside the metal. Notice that this is quite different from a dielectric, where the form of the electrostatic potential is unchanged but the magnitude is reduced by the dielectric constant ϵ .

General remarks on screening

The charge density we introduce into the solid we will call $\rho_{ext}(\mathbf{r}) = |e|n_{ext}(\mathbf{r})$. In vacuum it would produce a potential v_{ext} that satisifies Poisson's equation

$$\nabla^2 v_{ext}(\mathbf{r}) = \frac{e^2}{\epsilon_0} n_{ext}(\mathbf{r}) \quad . \tag{8.1}$$

Once the electrons in the solid relax to accommodate the new potential, the total charge density will consist of the external charge and the induced electron charge density $n_{ind}(\mathbf{r})$, viz

$$\rho_{tot}(\mathbf{r}) = \rho_{ext}(\mathbf{r}) + \rho_{ind}(\mathbf{r}) = e(n_{ext}(\mathbf{r}) - n_{\mathbf{r}}(\mathbf{r})) \quad , \tag{8.2}$$

which generates the actual potential v_{tot} seen by the electrons,

$$\nabla^2 v_{tot}(\mathbf{r}) = \frac{e^2}{\epsilon_0} (n_{ext}(\mathbf{r}) - n_{ind}(\mathbf{r})) \quad . \tag{8.3}$$

The net effect will be that the total potential seen by an individual electron in the Schrödinger equation is less than the external potential. In general, this phenomenon is incorporated into electromagnetic theory through the dielectric function ϵ , and usually one assumes that the total potential and the induced potential are linearly related (linearity being an assumption, for the moment). The dielectric function relates the electric displacement D to the electric field E, in the form

$$\mathbf{D}(\mathbf{r}) = \epsilon_0 \int d\mathbf{r}' \ \epsilon(\mathbf{r}, \mathbf{r}') \mathbf{E}(\mathbf{r}') \quad . \tag{8.4}$$

Because **D** is generated by "free" charges (i.e. ρ_{ext}) and **E** by the total charge (i.e. ρ_{tot}) Eq. (8.4) can be rewritten as a relationship between the potentials generated by those charge distributions:

$$v_{ext}(\mathbf{r}) = \int d\mathbf{r}' \ \epsilon(\mathbf{r}, \mathbf{r}') v_{tot}(\mathbf{r}') \quad . \tag{8.5}$$

In a spatially uniform electron gas, ϵ can depend only on the separation between coordinates, i.e. $\epsilon(\mathbf{r}, \mathbf{r}') = \epsilon(\mathbf{r} - \mathbf{r}')$, so that Eq. (8.5) becomes a convolution in real space – better written of course in Fourier (momentum) space as¹

$$v_{ext}(\mathbf{q}) = \epsilon(\mathbf{q})v_{tot}(\mathbf{q}) \quad . \tag{8.6}$$

The job of a microscopic model is to calculate the induced charge density (using the Schrodinger equation) produced by the external potential, i.e. $n_{ind}([v_{ext}(\mathbf{r})])$; again assuming it is linear (expected to be the case if v_{ext} is small enough, this will be a calculation of what is usually called the *linear* susceptibility χ , defined by

$$n_{ind}(\mathbf{q}) = \chi(\mathbf{q})v_{ext}(\mathbf{q}) \quad . \tag{8.7}$$

Looking back to the definitions of the potential in terms of the density, the relationship to the dielectric function is

$$\frac{1}{\epsilon(\mathbf{q})} = 1 - \frac{4\pi e^2}{q^2} \chi(\mathbf{q}) \quad . \tag{8.8}$$

Thomas-Fermi approximation

The Thomas-Fermi theory of screening starts with the Hartree approximation to the Schrödinger equation. The Hartree approximation is to replace the manybody pairwise interaction between the electrons by a set of interactions between a single electron and the charge density made up from all the other electrons, i.e. by a one-body potential for the i^{th} electron

$$U_{coul}(\mathbf{r}) = -e \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{\mathbf{r} - \mathbf{r}'} = e^2 \sum_{j \neq i} \int d\mathbf{r}' \frac{|\psi_j(\mathbf{r}')|^2}{\mathbf{r} - \mathbf{r}'} , \qquad (8.9)$$

 $^{^1\}mathrm{Here}$ we use the same symbol for a function and its Fourier transform, distinguishing the two by their arguments



Figure 8.1: Thomas-Fermi approximation

where the summation is over all the occupied states ψ_i . This clearly takes into account the averaged effect of the Coulomb repulsion due to all the other electrons. This introduces enormous simplicity, because instead of needing to solve an N-body problem, we have a (self-consistent) one-body problem. It contains a lot of important physics, and turns out to be an approximation that is good provided the electron density is high enough. We will discuss better theories later, in the special topic of the electron gas.

We shall treat the case of "jellium", where the ionic potential is spread out uniformly to neutralise the electron liquid. We wish to calculate the density induced by an external potential $n_{ind}([v_{ext}(\mathbf{r})])$.

Jellium. The potential in the problem is the *total* potential (external plus induced) produced by the added charge and by the non-uniform screening cloud (see Fig. 8.1)

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r}) + v_{tot}(\mathbf{r})\psi(\mathbf{r}) = E\psi(\mathbf{r}) \quad . \tag{8.10}$$

Slowly varying potential. Assume that the induced potential is slowly varying enough that the energy eigenvalues of Eq. (8.10) are still indexed by momentum, but just shifted by the potential as a function of position:

$$E(\mathbf{k}, \mathbf{r}) = \frac{\hbar^2 k^2}{2m} + v_{tot}(\mathbf{r}) \quad . \tag{8.11}$$

This makes only sense in terms of wavepackets, but provided the potential varies slowly enough on the scale of the Fermi wavelength $2\pi/k_F$, this approximation is reasonable.

Constant chemical potential. Keeping the electron states filled up to a constant energy μ requires that we adjust the local k_F such that

$$\mu = \frac{\hbar^2 k_F(\mathbf{r})^2}{2m} + v_{tot}(\mathbf{r}) = E_F(\mathbf{r}) + v_{tot}(\mathbf{r}) \quad , \tag{8.12}$$

where we have adopted the usual convention of measuring E_F from the bottom of the parabolic band².

Local density approximation. We assume that k_F just depends on the local density

$$k_F(\mathbf{r}) = (3\pi^2 n(\mathbf{r}))^{1/3}$$
 (8.13)

Remember that

$$v_{tot}(\mathbf{r}) = v_{ext}(\mathbf{r}) + v_{ind}(\mathbf{r}) = v_{ext}(\mathbf{r}) + \int d\mathbf{r}' \frac{e^2}{|\mathbf{r} - \mathbf{r}'|} n(\mathbf{r}')$$
(8.14)

Thomas-Fermi equation. Substituting into Eq. (8.12):

$$\frac{\hbar^2}{2m} (3\pi^2)^{2/3} n^{2/3}(\mathbf{r}) + \int d\mathbf{r}' \frac{e^2}{|\mathbf{r} - \mathbf{r}'|} n(\mathbf{r}') = \mu - v_{ext}(\mathbf{r}) .$$
(8.15)

Linearised Thomas-Fermi. When the added potential v_{ext} is small, the density cannot differ very much from the density n_o of the system without the potential so we may then *linearise*:

$$\delta n(\mathbf{r}) = n(\mathbf{r}) - n_o , \qquad (8.16)$$

where δn is the induced charge density. To linear order

$$\frac{\hbar^2}{3m} (3\pi^2)^{2/3} n_o^{-1/3} \delta n(\mathbf{r}) + \int d\mathbf{r}' \frac{e^2}{|\mathbf{r} - \mathbf{r}'|} \delta n(\mathbf{r}') = -v_{ext}(\mathbf{r})$$
(8.17)

Density response. This is solved by Fourier transforms.

$$\delta n(\mathbf{q}) = -\frac{v_{ext}(\mathbf{q})}{\frac{4\pi e^2}{q^2} \left[1 + \frac{q^2}{q_{TF}^2}\right]},$$
(8.18)

where the Thomas-Fermi wavevector (inverse of the screeing length) is

$$q_{TF}^2 = \frac{4}{\pi} \frac{me^2}{\hbar^2} k_F = \frac{4}{\pi} \frac{k_F}{a_B} = \left(\frac{2.95}{r_z^{\frac{1}{2}}} \mathring{A}^{-1}\right)^2.$$
(8.19)

Screening. $\epsilon_{TF} \propto q^{-2}$ at small q (long distances), so the long range part of the Coulomb potential (also $\propto 1/q^2$) is *exactly* cancelled. In real space, if $v_{ext} = Q/r$ is Coulombic (long range), $V(r) = (Q/r)e^{-q_{TF}r}$ is a short-range Yukawa, or screened potential³. In a typical metal, r_s is in the range 2 - 6, and so potentials are screened over a distance comparable to the interparticle spacing; the electron gas is highly effective in shielding external charges.

84

²One is often sloppy about using E_F and μ interchangeably; here is a place to take care

³This form is originally due to P.Debye and E.Hückel, *Zeitschrift für Physik* **24**, 185, (1923) and was derived for the theory of electrolytes; it appears also in particle theory under the name of the Yukawa potential; the physics in all cases is identical

8.2. PLASMA OSCILLATIONS

Dielectric function. Eq. (8.18) gives us the definition of the density response function χ , and hence we have calculated the static dielectric function in the Thomas-Fermi approximation

$$\epsilon^{TF}(q) = 1 + \frac{q_{TF}^2}{q^2} \quad ,$$
 (8.20)

or equivalently

$$v_{tot}(q) = v_{ext}(q) \frac{q^2}{q^2 + q_{TF}^2} .$$
(8.21)

Qu.8.20 Thomas-Fermi screening

Check the formulae in Eq. (8.18) and Eq. (8.19). Suppose that the potential $v_{ext} = Q/r$, show that the induced charge density is then of the form

$$\delta n(r) \propto \frac{e^{-r/\xi}}{r}$$

and identify the screening length ξ .

8.2 Plasma oscillations

If you treat the electrons as a classical charged fluid, and displace them relative to the fixed ionic background (jellium again) by an amount $\mathbf{u}(\mathbf{r})$:

$$\mathbf{u} = \mathbf{u}_o e^{i(\mathbf{q} \cdot \mathbf{r} - \omega t)} \quad \mathbf{u}_o \parallel \mathbf{q} \tag{8.22}$$

This is a $longitudinal\ wave$ — displacement parallel to the wavevector.

The longitudinal displacement of the charge induces a polarisation

$$\mathbf{P} = -ne\mathbf{u} \tag{8.23}$$

and therefore an internal electric field in the sample which is

$$\epsilon_0 \mathbf{E} = -\mathbf{P} + \mathbf{D} \tag{8.24}$$

where \mathbf{D} is the displacement field, due to external forces only. \mathbf{D} is the applied field.

The equation of motion for the displacement is

$$m\ddot{\mathbf{u}} = -e\mathbf{E} = -\frac{ne^2}{\epsilon_0}\mathbf{u} - \frac{e}{\epsilon_0}\mathbf{D}$$
(8.25)

which after rearrangement and Fourier transformation gives

$$(-m\omega^2 + \frac{ne^2}{\epsilon_0})u = -\frac{e}{\epsilon_0}D \quad . \tag{8.26}$$

Plasmons. If D = 0, Eq. (8.26) describes modes of *free oscillation*: solutions of the form $u_o e^{i\omega_p t}$ oscillating at the plasma frequency $\omega_p = (ne^2/\epsilon_0 m)^{1/2}$ – which describes a massive mode⁴.

Dielectric function. The ratio of the internal electric field \mathbf{E} to the applied field \mathbf{D} gives the dielectric function - some straightforward reshuffling leads to

$$\epsilon_0 \left[1 - \frac{\omega_p^2}{\omega^2} \right] E_{\parallel} = D_{\parallel} \quad , \tag{8.27}$$

which defines the longitudinal dielectric function

$$\epsilon_{\parallel} = \left[1 - \frac{\omega_p^2}{\omega^2} \right] \quad . \tag{8.28}$$

Since D is generated (by Poisson's law) by the external potential, and E is generated by the screened potential, another way of writing the dielectric function is

$$V_{sc}(\mathbf{q},\omega) = \frac{V_{ext}(\mathbf{q},\omega)}{\epsilon(\mathbf{q},\omega)}$$
(8.29)

Defined this way, we have already estimated a formula for the static dielectric function in Eq. (8.20).

This classical discussion does not generate any dispersion for the plasmon, i.s. the plasma frequency is found to be q-independent. It turns out that the classical theory is *exact* for $\mathbf{q} \to 0$, but there are quantum transitions that are entirely missed at short wavevectors.

Since ϵ measures the charge response of a solid, then plasmons are generated by any charged probe. The classic experiment to observe plasmons is *Electron Energy Loss Spectroscopy* (EELS), where a high energy electron is sent into the sample and the energy loss monitored. As in any driven oscillator, energy is dissipated at or near the resonant frequency (with the width in frequency depending on the damping of the oscillator). An EELS spectrum will therefore have a peak near the plasma frequency.

8.3 Optical conductivity of metals

The equation of motion we wrote down for the electrons in a solid Eq. (8.26) assumes that electrons are accelerated in an applied field, and do not suffer any damping — it would predict the conductivity of a metal to be infinite. This is fixed up in a *Drude model* by adding a phenomenological damping term, that represents the effects of scattering. With the additional term, the modified equation of motion becomes

$$m\ddot{\mathbf{u}} + \gamma\dot{\mathbf{u}} = -e\mathbf{E} = -\frac{ne^2}{\epsilon_0}\mathbf{u} - \frac{e}{\epsilon_0}\mathbf{D}$$
(8.30)

⁴In sensible CGS units, used in most texts $\omega_p^2 = 4\pi n e^2/m$.



Figure 8.2: Generic diagram of an inelastic scattering experiment. The incident particle - in this case an electron is scattered to a final state of different energy and momentum. By comparing the incident and scattered spectra, one deduces the *energy loss spectrum* of the internal collective excitations in the medium. For high energy electrons - typically used in an EELS experiment - the momentum loss (q - k) is small.

which after Fourier transformation gives

$$(-m\omega^2 - i\gamma\omega + \frac{ne^2}{\epsilon_0})u = -eD \quad . \tag{8.31}$$

Clearly we can rework the previous analysis by replacing everywhere $m\omega^2$ by $m\omega^2 + i\gamma\omega$. We will then arrive at a *complex* dielectric function

$$\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + i\omega/\tau} \quad , \tag{8.32}$$

where we have defined a relaxation rate

$$\frac{1}{\tau} = \frac{\gamma}{m} \quad . \tag{8.33}$$

This is one way of expressing the result. An alternative expression is not to relate the displacement field to the electric field, but instead to calculate the current density

$$j = -ne\dot{u} = i\omega neu \tag{8.34}$$

induced by the applied electric field. After a few algebraic manipulations, we get

$$j(\omega) = \frac{\epsilon_0 \omega_p^2}{(1/\tau - i\omega)} E(\omega) \quad . \tag{8.35}$$



Figure 8.3: . Electron energy loss spectrum for Ge and Si (dashed lines) compared to values of $Im(1/\epsilon)$ extracted directly from measurements of the optical conductivity. [From H.R.Philipp and H.Ehrenreich, Physical Review **129**, 1550 (1963)

This instead expresses the result as a complex conductivity

$$\sigma(\omega) = \frac{\epsilon_0 \omega_p^2}{(1/\tau - i\omega)} \quad . \tag{8.36}$$

The two results Eq. (8.32) and Eq. (8.36) are completely equivalent: a relation between D and E implies a relation between j and E. One may translate between the two formulae by

$$\epsilon(\omega) = 1 + i \frac{\sigma(\omega)}{\epsilon_0 \omega} \quad . \tag{8.37}$$

The real part of the conductivity is proportional to the imaginary part of the dielectric function.

From Eq. (8.36), one can see that the theory gives rise to a familiar d.c. conductivity

$$\sigma(0) = \epsilon_0 \omega_p^2 \tau = \frac{ne^2 \tau}{m} \quad , \tag{8.38}$$

so that τ has a simple interpretation as the mean free time between collisions. At frequencies larger than $1/\tau$ the conductivity rapidly falls off:

$$\Re\sigma(\omega) = \frac{\sigma(0)}{1 + \omega^2 \tau^2} \quad . \tag{8.39}$$

Consequently, the dielectric function near q = 0 can be extracted directly from measurements of the ac conductivity $\sigma(\omega)$ by optical absorption of light. A comparison between optical measurements and electron energy loss measurements is shown in Fig. 8.3.

Qu.8.21 Reflectivity of metals

The phase velocity of light in a conducting medium is the speed of light divided by the complex dielectric constant $N(\omega) = \epsilon(\omega)^{1/2}$ where we may use for ϵ the Drude result

$$\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + i\omega/\tau} \quad . \tag{8.40}$$

In a "good" metal, we have $1/\tau \ll \omega_p$.

Sketch curves of: (a) $\Re \sigma(\omega)$, (b) $\Re \epsilon(\omega)$, (c) $\Re N(\omega)$, (d) $\Im N(\omega)$.

You may wish to choose a parameter $\omega_p \tau \approx 100$ and let a computer do it. But make sure you understand the answer. In particular, notice that:

(a) For $\omega \ll 1/\tau$, ϵ is large and pure imaginary, so that $|N| \gg 1$ and has roughly equal real and imaginary parts,

(b) For $1/\tau \ll \omega \ll \omega_p$, ϵ is real and negative, so that N is imaginary, (c) For $\omega > \omega_p$, ϵ is positive, and N is real.

Consider a light wave with the electric field polarised in the x-direction at normal incidence from the vacuum on a good Drude metal (with $1/\tau \ll \omega_p$) occupying the region z > 0. In the vacuum (z < 0), the incident E_1 and reflected E_2 waves give rise to a field

$$E_x = E_1 \exp(i\omega[z/c - t]) + E_2 \exp(-i\omega[z/c + t]) \quad , \tag{8.41}$$

whereas in the medium, the electric field is

$$E_x = E_0 \exp(i\omega[N(\omega)z/c - t]) \quad . \tag{8.42}$$

Matching the electric and magnetic fields on the boundary, show that

$$E_0 = E_1 + E_2 \quad , \tag{8.43}$$

$$NE_0 = E_1 - E_2 , (8.44)$$

and hence show that the reflection coefficient satisfies

$$R = \left|\frac{E_2}{E_1}\right|^2 = \left|\frac{1-N}{1+N}\right|^2 \quad .$$
 (8.45)

Using the Drude formula above, show that

$$R \approx 1 - 2\left(\frac{\omega}{2\pi\sigma(0)}\right)^{1/2} \text{ for } \omega \ll 1/\tau$$
 (8.46)

$$\approx 1 - \frac{2}{\omega_p \tau} \qquad \text{for } 1/\tau \ll \omega \ll \omega_p$$
 (8.47)

$$\approx \frac{1}{16} \left(\frac{\omega_p}{\omega}\right)^4 \qquad \text{for } \omega_p \ll \omega \qquad (8.48)$$

and sketch the reflectivity $R(\omega)$.

To get the first two of these results with the minimum of fuss, you may find it helpful to expand in 1/N, viz.

$$R = \frac{(1 - \frac{1}{N})(1 - \frac{1}{N^*})}{(1 + \frac{1}{N})(1 + \frac{1}{N^*})} \approx 1 - 4\Re(1/N)$$
(8.49)

Qu.8.22 Optical properties of solids

Discuss why, at optical frequencies, glass is transparent, and silver is shiny, while graphite appears black, and powdered sugar is white.

Qu.8.23 Metals and insulators

Explain the differences between a metal and an insulator. Your discussion should include reference to: single particle properties; screening of the Coulomb interaction; optical properties; and electrical and thermal properties.

Chapter 9

Semiclassical model of electron dynamics

9.1 Wavepackets and equations of motion

We now want to discuss the dynamics of electrons in energy bands. Because the band structure is dispersive, we should treat particles as wave-packets. The band energy $\epsilon(\mathbf{k})$ is the frequency associated with the phase rotation of the wavefunction, $\psi_{\mathbf{k}}e^{-i\epsilon(\mathbf{k})t/\hbar}$, but for the motion of a wave packet in a dispersive band, we should use the group velocity, $d\omega/dk$, or as a vector

$$\dot{\mathbf{r}} = \mathbf{v}_g = \hbar^{-1} \nabla_{\mathbf{k}} \epsilon(\mathbf{k}) \quad , \tag{9.1}$$

where \mathbf{r} is the position of the wavepacket. All the effects of the interaction with the lattice are contained in the dispersion $\epsilon(\mathbf{k})$.

If a force F is applied to a particle, the rate of doing work on the particle is

$$\frac{d\epsilon_{\mathbf{k}}}{dt} = \frac{d\epsilon}{dk}\frac{dk}{dt} = Fv_g \tag{9.2}$$

which leads to the key relation

$$\hbar \frac{d\mathbf{k}}{dt} = \mathbf{F} = -e(\mathbf{E} + \mathbf{v} \wedge \mathbf{B}) = -e(\mathbf{E} + \hbar^{-1} \nabla_{\mathbf{k}} \epsilon(\mathbf{k}) \wedge \mathbf{B})$$
(9.3)

where we have introduced electric E and magnetic B fields.

The effect of an electric field is to shift the crystal momentum in the direction of the field, whereas the effect of a magnetic field is conservative - the motion in k-space is normal to the gradient of the energy. Thus a magnetic field causes an electron to move on a line of constant energy, in a plane perpendicular to the magnetic field. This property is the basis of magnetic techniques to measure the fermi surface of metals.



Figure 9.1: Band energy E(k) (solid line) and group velocity v(k) dashed line in a simple 1D band. A wavepacket progressing its crystal momentum according to Eq. (9.4) accelerates as k increases from zero, and then slows and reverses direction as k approaches the zone boundary.

Bloch oscillations

Suppose we have a one-dimensional electron band, such as shown in in Fig. 9.1. The group velocity is also shown — note that it reaches maximum size about half way to the zone boundary, and then decreases to zero at the zone boundary. If an electron in this band were subject to a constant electric field, we get

$$k(t) = k(0) - \frac{eEt}{\hbar} \quad , \tag{9.4}$$

so that the wave packet of electrons oscillates up and down the energy surface. It we start from the minimum of the band, then the group velocity grows linearly in time as for a free electron accelerating (though with a mass different from the free electron mass). However, on approaching the zone boundary, the group velocity slows - the acceleration of the particle is *opposite* to the applied force. What is actually happening is buried within the semiclassical model via the dispersion $\epsilon(k)$: as the wavepacket approaches the Brillouin zone boundary, real momentum (not crystal momentum k) is transferred to the lattice, so that on reaching the zone boundary the particle is Bragg-reflected.

Thus a DC electric field may be used - in principle - to generate an AC electrical current. All attempts to observe these *Bloch oscillations* in conventional solids has so far failed. The reason is that in practice it is impossible to have wavepackets reach such large values of momentum as π/a due to scattering from impurities and phonons in the solid. We will incorporate scattering processes in the theory in a moment.

It turns out however, that one can make *artificial* periodic potentials in a semiconductor superlattice. The details of this process will be discussed later, but for our purposes the net effect is to produce a square well potential that is periodic with a periodicity that can be much longer than the atom spacing. The corresponding momentum at the zone boundary is now much smaller, so the



Figure 9.2: . Schematic diagram of energy versus space of the conduction and electron bands in a periodic heterostructure lattice. The tilting is produced by the applied electric field. The levels shown form what is called a *Wannier-Stark* ladder for electron wavepackets made by excitation from the valence band in one quantum well, either vertically (n=0) or to neighbouring ($n = \pm 1$) or next-neighbouring $n = \pm 2$ wells of the electron lattice. In the experiment, electrons (and holes) are excited optically by a short-pulse laser whose frequency is just above the band gap of the semiconductor (i.e. a few $\times 10^{15}$ Hz). The electrical radiation (on a time scale of picoseconds) is monitored as a function of time and for different DC electrical biases, shown on the left panel. The spectral content is then determined by taking a fourier transform of the wavepackets (right panel); at large negative voltages one sees a peak at a frequency that increases with increasing bias. The device is not symmetric, and therefore has an offset voltage of about -2.4 V before the Bloch oscillation regime is reached. [From Waschke et al. Physical Review Letters **70**, 3319 (1993).]

wavepacket does not have to be excited to such high velocities. The signature of the Bloch oscillations is microwave radiation produced by the oscillating charge - at a frequency that is proportional to the DC electrical field.

Qu.9.24 Bloch oscillations

The valence electron energy band in a one-dimensional periodic structure is given by

$$\epsilon(k) = C\sin^2\left(\frac{ka}{2}\right) \tag{9.5}$$

Sketch the variation of the group velocity and the effective mass in the first Brillouin zone.

Show that in a uniform applied electric field, E, the motion in real space is periodic with amplitude $\frac{C}{2eE}$ and angular frequency $\frac{aeE}{\hbar}$.

The sample shown in Fig. 9.2 had an overall thickness of 1 μm . Using the measured frequency shift with bias $d\omega_{Bloch}/dV$ at the most negative voltages estimate the period of the superlattice potential.

Estimate the frequency and real space amplitude of Bloch oscillations produced by applying an electric field of $100 Vm^{-1}$ to electrons in a perfect crystal of GaAs. Is this observable in a real material?

Approximations and justification for the semiclassical model

A full justification of the semiclassical model is not straightforward and we will not go into that here. [See Kittel, Appendix E, and for a more formal treatment J. Zak, Physical Review **168** 686 (1968)]

- Note that at least the semiclassical picture takes note of the fact that the Bloch states are stationary eigenstates of the full periodic potential of the lattice, and so there are no collisions with the ions.
- We must be actually describing the motion of a wavepacket

$$\psi_n(\mathbf{r}, t) = \sum_{\mathbf{k}'} g(\mathbf{k} - \mathbf{k}') \psi_{n\mathbf{k}'}(\mathbf{r}, t) \exp\left[-i\epsilon_n(\mathbf{k}')t/\hbar\right] \text{ where } g(k) \to 0 \text{ if } |\mathbf{k}| > \Delta k$$
(9.6)

The wavepacket is described by a function $g(\mathbf{k})$ that is sharply peaked, of width Δk , say. Clearly $\Delta k \ll 1/a$, with a the lattice constant (otherwise the packet will disperse strongly).

- The size of the packet in real space is therefore $\Delta R \approx 1/\Delta k$. Consequently the semiclassical model can only be used to describe the response to fields that vary *slowly* in space, on a scale much larger than the lattice constant.
- n, the band index, is assumed to be a good quantum number. Clearly if the lattice potential were tiny, we would expect to return to free electrons, and be able to accelerate particles to high energies and make transitions between bands. Rather naturally the constraint is that the characteristic field energies be small in comparison to the band gap E_{gap} : they are in fact

$$eEa \ll E_{gap}^2 / E_F \tag{9.7}$$

with E_F the characteristic fermi energy, or overall bandwidth. The electric fields in a metal rarely exceed 1 Vm^-1 , when the LHS of this inequality is about 10⁻¹⁰ eV; not in danger.

• The corresponding constraint on magnetic fields is

$$\hbar\omega_c \ll E_{gap}^2/E_F \tag{9.8}$$

with $\omega_c = eB/mc$ the cyclotron frequency. This corresponds to about 10^{-2} eV in a field of 1 Tesla, so that strong magnetic fields indeed may cause transitions between bands, a process of magnetic breakdown.

• The last condition is that of course that the frequency of the fields must be much smaller than the transition energies between levels, i.e. $\hbar\omega \ll E_{gap}$

9.2 Electrons and holes in semiconductors

An immediate consequence of this picture is that filled bands are inert. If all the electrons states in a zone are occupied, then the total current is got by integrating the group velocity over the whole zone; but the group velocity is the gradient of a periodic function; so this integral yields zero. Indeed all insulating solid elements have either even valence, or a lattice containing an even number of atoms in the basis, and therefore filled bands.

It is of interest to consider what happens to a filled band with one electron removed. This can be created by absorption of a photon whose energy exceeds the energy gap of a semiconductor, to make a transition of an electron from the valence band into the conduction band [See Fig. 9.3]. The removal of an electron from a filled band leaves a *hole*, which in fact can be viewed as a fermionic particle with distinct properties.

Hole momentum.

$$\mathbf{k}_h = -\mathbf{k}_e \tag{9.9}$$

This can be seen from the optical absorption experiment. The light produces a (nearly) vertical transition and gives no momentum to the electron hole pair. Since the initial state is a filled band with total momentum zero, Eq. (9.9) follows.

Hole energy.

$$\epsilon_h(\mathbf{k}_h) = -\epsilon_e(\mathbf{k}_e) \tag{9.10}$$

This sign is needed because (measuring energies from the top of the band) removing an electron of *lower* energy requires more work.

Hole velocity. A combination of the first two rules then gives

$$\mathbf{v}_{\mathbf{h}} = \hbar^{-1} \nabla_{\mathbf{k}_h} \epsilon_h(\mathbf{k}_h) = \hbar^{-1} \nabla_{\mathbf{k}_e} \epsilon_e(\mathbf{k}_e) = \mathbf{v}_e \tag{9.11}$$

Effective mass. The dispersion at the bottom (top) of the bands is parabolic, and therefore can be approximated as

$$\epsilon = \epsilon_0 + \frac{\hbar^2 k^2}{2m^*} \tag{9.12}$$



Figure 9.3: Absorption of a photon creates an electron-hole pair, with an energy $\epsilon_e + \epsilon_h + 2E_{gap}$ but adds negligible momentum to the system. Hence the hole momentum is the negative of the momentum of the empty electronic state, and its energy is positive (measured conventionally from the top of the band).

defining an effective mass m^* . We have

$$m_h^* = -m_e^* \tag{9.13}$$

so the hole mass is positive at the top of the electron band.

Hole charge. The effective charge of a hole is positive, as can be seen by taking the equation of motion for the electron

$$\hbar \frac{d\mathbf{k}_e}{dt} = -e(\mathbf{E} + \mathbf{v_e} \wedge \mathbf{B}) \tag{9.14}$$

and making the replacement $k_e \rightarrow -k_h$ and $v_e \rightarrow v_h$, giving

$$\hbar \frac{d\mathbf{k}_h}{dt} = e(\mathbf{E} + \mathbf{v}_h \wedge \mathbf{B}) \tag{9.15}$$

The same result comes from noticing that the current carried by the hole ev_h must be the same as the (missing) current (not) carried by the empty electron state.

Qu.9.25 Hole current and electron current An applied DC electric field points along the positive x-axis of a semiconductor containing both an electron in the conduction band and a hole in the valance band.

(a) In which direction is the initial velocity of the electron? What is the direction of the electrical current carried by the electron?

(b) In which direction is the initial velocity of the hole? What is the direction of the electrical current carried by the hole?

Qu.9.26 Cyclotron resonance A DC magnetic field B points in the z-direction. An electron sits near the bottom of a parabolic band with effective mass m^* . Show that the particle orbits with an angular velocity

$$\omega_c = \frac{eB}{m^*} \tag{9.16}$$

Is the sense of the orbit the same, or different, for a hole?

9.3 Scattering and electrical conductivity of metals

We will incorporate scattering into our semiclassical picture in the crudest possible way, though frequently quite adequate at a phenomenological level, by introducing a scattering rate τ . The equation of motion becomes

$$\hbar \frac{d\mathbf{k}}{dt} = \mathbf{F} - \hbar \frac{\mathbf{k}}{\tau} = -e(\mathbf{E} + \mathbf{v} \wedge \mathbf{B}) - \hbar \frac{\mathbf{k}}{\tau}$$
(9.17)

so that the collisions are introduced as a frictional force. Let us now apply this simple theory to describe the electrical conductivity of a metal.

We set B = 0, and consider first the fermi sea at rest. There are equal numbers of states occupied of positive and negative momentum, and hence there is no current in the absence of an applied field. In a small electric field E, after an initial transient, the result of Eq. (9.17) will be that the momenta of all particles in the fermi sea will be shifted by a small amount

$$\delta \mathbf{k} = -eE\tau/\hbar \tag{9.18}$$

For a parabolic band, the incremental velocity is

$$\delta \mathbf{v} = \delta \mathbf{k} / m^* \tag{9.19}$$



Figure 9.4: The net current from the fermi sea at rest is zero, but a shift by a small momentum leads to a net imbalance.

and therefore a net current density

$$\mathbf{j} = -ne\delta \mathbf{v} = \frac{ne^2\tau}{m^*}\mathbf{E}$$
(9.20)

which is the familiar Ohm's law.

The conductivity can also be written in terms of the mobility $\mu = e\tau/m^*$

$$\sigma = ne\mu = \frac{ne^2\tau}{m^*} \tag{9.21}$$

Almost invariably, $\delta k \ll k_F$, so the shift in the Fermi sea is very small and the net current is entirely produced by the small imbalance near the Fermi surface. Just as for the specific heat and paramagnetic suspetibility, it is only the properties of electrons near the fermi surface that determine the conductivity of a metal.

Thermal conductivity of metals

Particles with velocity v, mean free path l and specific heat C are expected to yield a thermal conductivity K = Cvl/3. For a free fermi gas, we get the correct answer from this formula by using the electronic specific heat, the characteristic carrier velocity v_F , and the mean free path for carriers on the fermi surface $l = v_F \tau$. Hence, using the relationship between the Fermi velocity and the fFermi energy $E_F = mv_F^2/2$,

$$K_{el} = \frac{\pi^2}{2} \frac{nk_B^2 T}{E_F} \cdot v_F \cdot v_F \tau = \frac{\pi^2 nk_B^2 T \tau}{3m}$$
(9.22)

Almost invariably, the electronic thermal conductivity is bigger than that due to the lattice. K and σ are of course closely related, being both proportional to the scattering time and the density, as is natural. The ratio

$$\frac{K}{\sigma T} = \frac{\pi^2}{3} \left(\frac{k_B}{e}\right)^2 \tag{9.23}$$

is expected to be a constant, independent of material parameters. This proportionality is the *Wiedemann-Franz* law, which works strikingly well for simple metals.

9.4 Transport in electrical and magnetic fields

Returning to Eq. (9.17), we now study the electrical transport in a transverse magnetic field, i.e. a static magnetic field **B** in the \hat{z} direction, and static currents and electrical fields in the x - y plane.

The equations of motion for an electron of charge -e are now¹.

$$m(\partial_t + \tau^{-1})v_x = -e(E_x + Bv_y)$$

$$m(\partial_t + \tau^{-1})v_y = -e(E_y - Bv_x)$$

$$m(\partial_t + \tau^{-1})v_z = -eE_z$$
(9.24)

In steady state , we set the time derivatives $\partial_t = d/dt = 0$, and get the three components of the drift velocity

$$v_x = -\frac{e\tau}{m}E_x - \beta v_y$$

$$v_y = -\frac{e\tau}{m}E_y + \beta v_x$$

$$v_y = -\frac{e\tau}{m}E_z$$

(9.25)

with the dimensionless parameter $\beta = \frac{eB}{m}\tau = \omega_c\tau = \mu B$ the product of the cyclotron frequency and the scattering rate.

Hall effect

If a current $\mathbf{j} = e\mathbf{v}$ is flowing in a magnetic field \mathbf{B} , Eq. (9.17) leads to a force on the carrier $-e\mathbf{v} \wedge \mathbf{B}$, normal to the direction of flow. Since there is no flow in the normal direction, there must exist a counterbalancing electric field $\mathbf{E} = -\mathbf{v} \wedge \mathbf{B}$. This is the *Hall effect*.

Consider the rod-shaped geometry of Fig. 9.5. The current is forced by geometry to flow only in the x-direction, and so $v_y = 0$, which gives

$$v_x = -\frac{e\tau}{m} E_x \quad , \tag{9.26}$$

and

$$E_y = -\beta E_x \tag{9.27}$$

It turns out that for high mobility materials, and large magnetic fields, it is not hard to reach large values of $\beta \gg 1$, so that the electric fields are largely normal to the electrical currents.

¹We define e to be a positive constant.



Figure 9.5: The upper figure shows the geometry of a *Hall bar*, with the current flowing uniformly in the x-direction, and the magnetic field in z. The lower figure shows the *steady state* electron flow (arrows) in a section normal to \hat{z} . When a voltage E_x is first applied, and E_y is not yet established, the electrons will deflect and move in the (downward) y-direction. The y-surfaces of the crystal then become charged, producing the field E_y which exactly cancels the Lorentz force $-ev_xB$.

The Hall coefficient is defined by

$$R_H = \frac{E_y}{j_x B} = -\frac{1}{ne} \tag{9.28}$$

Notice that it is *negative* for electrons, but importantly is independent of the effective mass, and grows with decreasing carrier density. One can straightforwardly check that for holes of charge +e the sign is positive.

The Hall effect is an important diagnostic for the density and type of carriers transporting the electrical current in a semiconductor. The simple picture of parabolic bands works quite well for alkali metals, where the predicted Hall coefficient is within a few percent of the expected value for a parabolic free electron band. But Be, Al, and In all have *positive* Hall coefficients - accounted for by a band-structure with hole pockets that dominates the Hall effect.

Qu.9.27 Static conductivity tensor

From Eq. (9.28), show that the electrical conductivity can be written as a tensor $\mathbf{j} = \sigma \cdot \mathbf{E}$, where the cartesion components are

$$\sigma = \frac{\sigma_o}{1+\beta^2} \begin{pmatrix} 1 & -\beta & 0\\ \beta & 1 & 0\\ 0 & 0 & 1+\beta^2 \end{pmatrix}$$
(9.29)

Here $\beta = \omega_c \tau$ and $\sigma_o = n e^2 \tau / m$.

In a high magnetic field ($\beta \gg 1$), show that $\sigma_{yx} = -\sigma_{xy} = ne/B$.

102

Chapter 10

Semiconductors

10.1 Semiconductor band structure

Direct gap semiconductors

The band structure near k = 0 of a diamond-structure (Si, Ge) or zincblendestructure (GaAs) semiconductor is shown in Fig. 10.1. The conduction band is a simple parabola, but the valence bands are more complex. The complexity arises because the symmetry of the valence bands is p - like and there are three degenerate bands (in cubic symmetry) at k = 0. At finite k they split into *light hole* and *heavy hole* bands, so called because of the difference in the electron masses. Additionally, there is a deeper lying band, split off by spinorbit interactions from the others. This is usually not important for thermally excited carriers.

The band masses are quite different from free electron masses, for example, in GaAs $m_e^* = 0.066$, $m_{lh}^* = 0.082$, $m^*hh = 0.17$ (in units of the free electron mass). The cubic symmetry of the crystal means that the bands are isotropic (to order k^2).

Qu.10.28 Band gaps and effective masses Using the onedimensional nearly-free electron model of Eq. (6.14), calculate the effective masses of electrons and hole states in terms of the band gap.

Is the data in the table below approximately consistent with your result?

Crystal	m_e^*	$E_{gap}/{ m eV}$
InSb	0.015	0.23
InAs	0.026	0.43
InP	0.073	1.42



Figure 10.1: Sketch of the valence bands of diamond or zincblende structure semiconductors near the Γ -point ($\mathbf{k} = (000)$). The lowest hole band - the spin-orbit split-off band - is lower by an energy Δ that is a few tenths of an eV and therefore not relevant for thermally excited carriers at room temperature and below. In III-V semiconductors the absolute minimum in the conduction band is at Γ ; in *Si* and *Ge* the absolute minimum of the conduction band is elsewhere in the zone.

Indirect gap semiconductors

As we remarked earlier, the conduction bands of Si and Ge do not have their minima at the Γ -point, but far out in the zone.

The conduction band minima of Ge are at the eight equivalent L-points $2\pi/a(0.5\ 0.5\ 0.5)$, on the surface of the Brillouin zone. Here the band edges have a spheroidal energy surface, and are not isotropic as near the centre of the zone. In Ge, the longitudinal mass – along (111) – is $m_l = 1.59 m$, much larger than the tranverse mass $m_t = 0.082 m$.

In Si the conduction band minima are along the six (100) directions, close to the zone boundary at X $[2\pi/a(100)]$. The constant energy surfaces are ellipsoids, $m_l = 0.92 m$, and $m_t = 0.19 m$.

Bands near the zone centre

The band structure near k = 0 is simple to analyse, because Bloch's theorem here requires the wave function to be the same in every unit cell. The wavefunctions have pure symmetry requirements imposed on them by the lattice. Since the zincblende structure has inversion symmetry about each atom the states can be separated into those that are even or odd about any lattice site (for our purposes s and p). Additionally, if we have the diamond lattice there is a further symmetry point midway between the two atoms in the unit cell; the states can also be classified as odd/even about this point. So we must have the following sets of states:



Figure 10.2: Valence and conduction bands near k=0 according to a tightbinding prescription. (a) is for diamond crystals, and (b) for zincblende; spinorbit splitting is neglected. (From Phillips.)

bonding (antibonding) $s_1 \pm s_2$ bonding (antibonding) $p_1 \pm p_2$

(the subscript refers to the two atoms in the cell) The p- states are triply degenerate (in the absence of spin-orbit coupling which we will come back to in a moment) because the crystal is cubic – this degeneracy is analogous to the orbital degeneracy of p_x, p_y and p_z states in an atom. We expect these states to be ordered in the way shown in Fig. 10.2.

In zincblende materials, the admixture of states will reflect the partial ionicity, as noted in the figure: if atom 2 is the more electronegative atom, $\gamma^{(\ell)} > 1$, but the bonding/antibonding character is essentially preserved.

Now let us include spin-orbit interactions, which will come with a term $\lambda \mathbf{L} \cdot \mathbf{S}$ in the Hamiltonian. This will split the *p*-states in Fig. 10.2, and we can easily see how. Without spin orbit, we had a sixfold-degenerate level altogether -3 (orbital) x 2 (spin). After coupling, the states must be eigenfunctions of the total angular momentum J, which can take on the values J = 3/2 (fourfold degenerate), and J = 1/2 (twofold degenerate). We end up with the bands shown in Fig. 10.1 and in particular with the characteristic "light" and "heavy" mass bands for holes.



Figure 10.3: Density of states for electrons and the fermi function determining the occupancy of the thermally excited states in an intrinsic semiconductor. The chemical potential lies mid-gap, and the temperature is assumed small in comparison to the gap.

10.2 Intrinsic carrier concentration

Semiconductors are materials where the energy gap is small enough that thermal excitation of carriers across the gap is important. Here we shall calculate the thermal *intrinsic* carrier concentration in a model semiconductor with parabolic electron and hole bands. The conduction and valence band dispersions are thus (see Fig. 10.3)

$$E_c(k) = E_c + \frac{h^2 k^2}{2m_e^*} ; \quad E_v(k) = E_v - \frac{h^2 k^2}{2m_h^*}$$
(10.1)

We shall need the densities of states for the conduction band

$$g_e(E) = \frac{1}{2\pi^2} \left(\frac{2m_e^*}{\hbar^2}\right)^2 (E - E_c)^{1/2}$$
(10.2)

and for the valence band

$$g_h(E) = \frac{1}{2\pi^2} \left(\frac{2m_h^*}{\hbar^2}\right)^2 (E_v - E)^{1/2}$$
(10.3)

10.2. INTRINSIC CARRIER CONCENTRATION

We can calculate the carrier density once the chemical potential μ is known. For electrons in the conduction band

$$n = \int_{E_c}^{\infty} dE \, g_e(E) f(E) \tag{10.4}$$

with f the fermi function

$$f(E) = \frac{1}{e^{(E-\mu)/(k_B T)} + 1} \approx e^{-(E-\mu)/(k_B T)}$$
(10.5)

with the latter approximation valid for small enough temperature. This gives

0.10

$$n \approx 2 \left(\frac{m_e^* k_B T}{2\pi\hbar^2}\right)^{3/2} e^{-\frac{E_c - \mu}{k_B T}}$$
(10.6)

A similar calculation determines the concentration of holes

$$p \approx 2 \left(\frac{m_h^* k_B T}{2\pi\hbar^2}\right)^{3/2} e^{-\frac{\mu - E_v}{k_B T}}$$
(10.7)

Qu.10.29 Hole statistics Show that $f_h(\epsilon) = 1 - f_e(\epsilon)$, where

$$f_e(\epsilon) = \frac{1}{e^{\beta(\epsilon-\mu)}+1}$$

is the fermi distribution function for electrons and

$$f_h(\epsilon) = \frac{1}{e^{\beta(\mu-\epsilon)}+1}$$

describes the distribution function for holes.

Eq. (10.4) and Eq. (10.7) give the concentration of electrons and holes at a temperature T, in terms of the chemical potential μ , as yet unknown. It is useful to notice that the product

$$np = 4\left(\frac{k_B T}{2\pi\hbar^2}\right)^3 (m_e^* m_h^*)^{3/2} e^{-\frac{E_{gap}}{k_B T}}$$
(10.8)

is independent of μ . In particular, we will be able to use this result when carriers are introduced to the system by impurities.

For an intrinsic semiconductor, the electron and hole densities are equal, and can be obtained my taking the square root of Eq. (10.8)

$$n_i = p_i = 2\left(\frac{k_B T}{2\pi\hbar^2}\right)^{3/2} (m_e^* m_h^*)^{3/4} e^{-\frac{E_{gap}}{2k_B T}}$$
(10.9)

and substituting back in either the equation for n (Eq. (10.4)) or p Eq. (10.7)) yields the chemical potential

$$\mu = \frac{1}{2}E_g + \frac{3}{4}k_B T \log(m_h/m_e) \quad . \tag{10.10}$$

The chemical potential thus sits mid gap at zero temperature, and shifts slightly away from that position if the carrier masses are different. Note that the activation energy to create intrinsic carriers (either electrons or holes) is always exactly *half* the optical energy gap.

10.3 Extrinsic semiconductors

Carriers can also be created in semiconductors by adding impurity atoms in a process called *doping*.

Donor levels. Consider the effect in a Si crystal of replacing a single atom by an As atom. As is a group V element and therefore provides 5 electrons in the place of the 4 of the Si it replaced. We now ask whether the added electrons stay tightly bound to this impurity.

Suppose one electron wanders away from the impurity site. It will of course see an attractive Coulomb force from the (now charged) As impurity, and the energy levels are then that of the Hydrogen atom. The Coulomb potential is however screened by the dielectric constant of Si ($\epsilon \approx 12$) so is much weaker than in free space. Also, the electron effective mass is smaller than the free electron mass, so the kinetic energy of an electron in a given momentum state is larger. The net effect is that the binding energy of the 1s impurity state is now

$$\Delta_d = \frac{e^4 m_e^*}{2(4\pi\epsilon\epsilon_o\hbar)^2} = \frac{m_e^*/m}{\epsilon^2} \times 13.6 \text{ eV}$$
(10.11)

which can be very small in comparison to the band gap, and often comparable or smaller than thermal energies. Such *donor* impurities easily donate electrons to the conduction band.

As donors in Si have an ionisation energy of 50 meV; donors in GaAs have an ionisation energy of about 6 meV, approximately 50 Kelvin.

Acceptor levels. A trivalent impurity (e.g. B) in Ge or Si can bind a hole — or *accept* an electron from the valence band. The accepted electron is used to complete the covalent bond with the neighbouring atoms (which however renders the site *negatively* charged). The hole left in the band is weakly bound to the acceptor site, and easily ionised. The calculation of the binding energy of holes is more complex than for electrons because of the degeneracy of the valence bands.

n- and p-type materials. Even for very low densities of impurities, since the donor or acceptor energies are so much smaller than the gap, impurities


Figure 10.4: . The left hand figure shows the effect of donor levels at an energy $E_d = E_c - \Delta_d$. The chemical potential will shift near the conduction band edge, increasing the electron density and decreasing the hole density in comparison to the intrinsic case, while the product of the two is nevertheless constant. The right hand figure shows the corresponding picture for acceptor levels at an energy $E_a = E_v + \Delta_a$.

in semiconductors are often the principal source of electrically active carriers. If donor atoms predominate, the carriers are predominantly electrons, and the material is said to be *n*-type. If holes are the dominant carrier type, the material is called *p*-type. Experimentally, these regimes may be distinguished by a measurement of the Hall effect, whose sign depends on the type of carrier.

Impurity ionisation. We just quote here the results for thermal ionisation of the carriers in simple limits. If there are no acceptors, then the carrier concentration at low temperatures is

$$n = (n_c N_d)^{1/2} e^{-\frac{\Delta_d}{2k_B T}}$$
(10.12)

where N_d is the donor density and the factor $n_c = 2(m_e^* k_B T / 2\pi \hbar^2)^{3/2}$ is the effective density of electron states within an energy $k_B T$ of the band edge. Notice again that the activation energy is *half* the binding energy.

Since Δ_d is often quite small, there may be a regime of temperature $\Delta_d < 2k_BT < E_g$ where the donors are all ionised, and $n \approx N_d$. At a still higher temperature, the *intrinsic* carrier generation process takes over.

If there are only acceptors and no donors, then a similar formula can be

obtained for holes by inspection. When both donors and acceptors exist, the problem is more complicated and not illuminating. However, many situations of practical importance for devices require such an analysis.

Lastly, note that by the law of mass action, np is a constant at fixed temperature, given by Eq. (10.8). An excess of electron donors over hole acceptors will therefore increase n and decrease p.

Qu.10.30 Ge Give a brief explanation of the concepts of *drift velocity, electron mobility,* and *effective mass,* as used in solid state physics.

A sample of Ge is doped so that the concentration of pentavalent donor impurities, N_d is $3 \times 10^{22} m^{-3}$, and that of trivalent acceptors, N_a , is $10^{22} m^{-3}$. Estimate the concentration of electrons in the conduction band and holes in the valence band at 300 K.

[The intrinsic carrier density of Ge at 300K is $2.4 \times 10^{19} m^{-3}$.]

Sketch a graph of the conductivity as a function of temperature you would expect to measure for this sample of Ge.

Qu.10.31 Impurity Bands

InSb has a dielectric constant $\epsilon=18$ and an effective mass for electrons $m^*=0.015m.$ Calculate the ionisation energy of a hydrogenic donor orbit.

At what density of donors do you expect to see the effects of overlaps between the orbits of adjacent impurities?

At low densities, donor levels are isolated, and if the temperature is so low that the probability of ionisation is very small, the system will be an insulator. But at higher density, the donor levels overlap to form an impurity band that can support metallic conduction.

Chapter 11

Semiconductor devices

We now consider the properties of inhomogeneous systems and devices. In this section we discuss the general properties of surfaces and interfaces between materials, and then the basic devices of bulk semiconductor physics. For bulk devices we will use the semiclassical approximation, treating electrons as classical particles obeying the Hamiltonian¹

$$H = E_n(\mathbf{k}) - e\phi(\mathbf{r}) \tag{11.1}$$

with the momentum $\mathbf{p} = \hbar \mathbf{k}$ and a spatially varying potential $\phi(\mathbf{r})$. The potential will arise from externally applied fields, from charges induced by doping, and from changes in the material composition. When we discuss narrow quantum wells, we shall need to modify this approximation to quantise the levels.

For an isolated solid in equilibrium, the energy difference between the chemical potential μ the vacuum level is the work function Φ . This is the energy required to remove an electron from the fermi level and place it in a state of zero kinetic energy in free space.

Two different isolated materials with different work functions will then have different chemical potentials. If these two materials are placed in contact, their chemical potentials must equalise, and this is accomplished by electron flow to the more electronegative material; this material becomes charged, its potential ϕ changes, and an overall balance will be established. But in general, there will be as a result internal, inhomogeneous electric fields.

11.1 Metal - semiconductor contact

The figure Fig. 11.1 is a schematic of this process for an ideal metal placed in contact with a semiconductor. We consider the more interesting case when the

¹We shall keep to the convention that e is a positive number, and therefore $-e\phi$ is the potential energy seen by electrons in an electrostatic potential ϕ

chemical potential of the (doped) semiconductor is above that of the metal.

Qu.11.32 Depletion layer

A full treatment of this problem requires the solution of the Poisson equation to determine the electric field distribution V(x) combined with the thermal carrier statistics to determine the occupancy of the states. At low temperature, when the boundary of the depletion regime may be assumed to be sharp, it is more straightforward.

A metal-semiconductor contact is made between a perfect conductor and a uniformly doped n-type semiconductor with a donor density N_d (see Fig. 11.1. Assume that the temperature is low eneough that the donor levels are completely filled or completely empty. By solving Poisson's equation, show that in the depletion region $0 < x < x_b$ the potential satisfies

$$\phi = \phi_b - \frac{N_d e}{2\epsilon\epsilon_0} x^2 \tag{11.2}$$

Estimate the depletion width for a semiconductor with $\epsilon = 12$, $e\phi_b = 0.5 \ eV$, and $N_d = 10^{22} \ m^{-3}$.

Qu.11.33 Ohmic contact

Sketch the equivalent diagrams to Fig. 11.1 for a metal-semiconductor junction in the case that $\phi_s > \phi_m$. Explain why such a junction acts as an ohmic contact.

Rectification by a metal contact

The barrier set up between the metal and insulator inhibits current flow. An electron from the metal must either tunnel through the barrier (at low temperatures) or may be thermally excited over it (thermionic emission). However, when a large enough external bies is applied, the junction may act as a rectifier Fig. 11.2. We will not analyse this in detail, as the more important case of a p-n diode is similar, and soon to follow.

11.2 p-n junction

A p-n junction is formed by inhomogeneous doping: a layer of n-type material (containing donors) is placed next to p-type material (with acceptors). A schematic layout is shown in Fig. 11.3. The behaviour can be understood by an extension of the discussion for the metal-semiconductor contact.



Figure 11.1: (a) When metal and semiconductor are not in contact they are in equilibrium with the vacuum level. We consider an n-type semiconductor, with the chemical potential lying close to the conduction band edge. (b). When the two are brought into contact, electrons leave the semiconductor and are transferred to the metal. This produces an electrical potential $\phi(x)$ which will eventually equilibrate so that the chemical potential is constant over the whole system. (The combined function $\mu + e\phi(x)$ is sometimes called the *electrochemical potential*.) (c) Shows the energy level diagram relative to the constant chemical potential. The semiconductor bands bend upwards, so that the donor levels near the interface are emptied of electrons - leaving a positively charged *depletion region*, and a *Schottky barrier* ϕ_b .



Figure 11.2: . Schematic picture of a Schottky diode. In the upper panel, applying a positive bias across the junction lowers the barrier for electrons to enter the metal, and can eventually tilt the electron bands so much that the barrier disappears. The current grows rapidly with positive bias. However, if the bias is negative, the depletion width grows and the current is little changed.

- Deep inside the n-doped (p-doped) regimes, the chemical potential must lie close to the donor (acceptor) levels.
- If we were instantaneously to place the n-type and p-type regions in contact, charge will flow because of the different chemical potentials.
- In doing so, the interface region becomes depleted of carriers, and the ionised donors (acceptors) now have positive (negative) charge (see Fig. 11.4.
- The electrostatic potential so generated shifts the energy levels of the donors (acceptors) down (up) and the chemical potential is equalised

The typical extent of the depletion region is between 10 nm and 1 μ m.

Rectification by a p-n junction

A p-n junction behaves as a *diode*, allowing current to flow much more readily in one direction from the opposite. A simple picture can be given as follows, with reference to the diagram in Fig. 11.5. Our sign convention is to apply an electrical bias where *positive* voltage V is applied to the *p*-type side of the junction.

Diffusion currents. In equilibrium with no external voltage bias, there is no net current flowing across the junction. But of course to maintain chemical equilibrium, there must be microscopic current flows across the barrier in both



Figure 11.3: Two equivalent ways of representing the energy levels in a p-n junction. (a) shows the energy levels, and includes the electrostatic potential in the electrochemical potential $\mu + e\phi(x)$. In (b) we recognise that the chemical potential is constant, and the effect of the potential ϕ is a shift in the energetic position of the energy levels $E_d(x) = E_d - e\phi(x)$, $E_a(x) = E_a - e\phi(x)$. When the shifted donor or acceptor levels pass through the chemical potential, these levels are *ionised*, and the carriers pass from one side of the barrier to the other, and annahilate. The impurity levels within the depletion region are now charged.



Figure 11.4: (a) Carrier densities and (b) charge densities near the depletion region of a p-n junction. When the temperature is low, the carrier density changes abruptly at the point where the chemical potential passes through the donor or acceptor level. Close to the barrier, the carriers are depleted, and here the system is now physically charged, with a charge density of $+eN_d$ on the n-type side, and $-eN_a$ on the p-type side. This dipole layer produces a potential $\phi(x)$ shown in (b). The potential itself self-consistently determines the charge flow and the width of the depletion region.



Figure 11.5: Operation of a p-n diode. (a) shows the potential across the junction, which is decreased in forward bias (corresponding to a positive voltage V applied to the p-type side). (b) Shows the carrier density for holes - large on the p-type side and a small minority carrier population on the n-type side. Carriers diffusing out of either of the two populations generate opposing currents that exactly cancel in equilibrium at zero bias. (c) shows the same for the electrons. Because the charge on the electron is negative the electrical currents shown have the same sign as for holes, though of course the particles move in the opposite direction. In forward bias, the recombination current (which is thermally activated) grows exponentially with bias, and the device readily passes current. In reverse bias, there is very little change, and the currents are small.

directions, that cancel on average. We consider them separately for holes and electrons.

Potential barrier. The depletion regime of the junction is a high-resistance in comparison to the n- or p-type doped semiconductors. Any potential applied across the device is dropped almost entirely across the depletion layer. The overall potential seen by a (positively charged) hole is therefore $\phi_b - V$, where ϕ_b is the barrier height at equilibrium.

Hole generation current. On the n-type side of the depletion regime, the majority carriers are electrons, but detailed balance (Eq. (10.8)) means that there will always be some small density of *minority* holes. Any minority carrier diffusing into the depletion regime will be swept down the potential into the p-type regime. This generates a current (from right to left, of holes, and therefore negative)²

$$-J_h^{gen} \tag{11.3}$$

It is not much dependent on the external bias V, because of the large inbuilt potential drop in the depletion regime.

Hole recombination current. The holes in the p-type regime have a small probability of being thermally excited up the potential hill. Since the temperature is low compared to the barrier, this current is activated,

$$J_b^{rec} \propto e^{-e(\phi_b - V)/k_B T} \quad . \tag{11.4}$$

Net current. We know that in equilibrium at zero bias the hole recombination current and generation currents must cancel. Then the total hole current takes the form

$$J_h = J_h^{gen} (e^{eV/k_B T} - 1) \quad . \tag{11.5}$$

Electrons. The same analysis appplies to electrons, except that the corresponding electron generation and recombination (number) currents flow in the opposite directions to their hole counterparts. But electrons are oppositely charged, so the electrical current density has the *same* form as Eq. (11.5).

Diode IV characteristic. The sum of the contributions of electrons and holes gives an asymmetrical form

$$I = I_{sat} \left(e^{eV/k_B T} - 1 \right) \tag{11.6}$$

where the saturation current I_{sat} is proportional to n_i^2 and therefore of the Arrhenius form e^{-E_g/k_BT} (see footnote 2).

² The magnitude can be estimated to be $(n_i^2/N_d)(L_p/\tau_p)$, where the first factor in brackets is the minority hole density in the p-type region (Eq. (10.9)) τ_p is the recombination time of a carrier, and L_p is the length that the hole will diffuse before it recombines with an electron



Figure 11.6: The left panel shows the theoretical I-V characteristic from Eq. (11.6). The right panel shows data from an early Ge p-n diode compared to the same model. [F.S. Goucher, G.N. Pearson, M.Sparkes, G.K.Teal and W. Shockley, Physical Review **81**, 637, (1951)]

Solar cell

If light shines on a p-n junction, without an external bias voltage, then each absorbed photon will create an electron-hole pair. If these carriers reach the junction, the built-in field will separate them - the potential gradient pulls electron and hole in opposite directions. The separation of the charges across the depletion layer adds an extra internal dipole to the system - like charging a capacitor - and therefore generates an overall electrical bias. The induced voltage is in the forward direction - because it is opposite in sign to the built-in potential.

This is the *photovoltaic effect*, which can deliver power to an external circuit. Large arrays of p-n junctions of Si are used to make solar panels, converting solar radiation to electrical power.

Light-emitting diode

The inverse process to the photovoltaic effect powers a light-emitting diode or LED. Here a current is injected into a p-n diode in a non-equilibrium situation where the electron and hole chemical potentials are kept to be different with a large bias eV. Electrons are injected from *n*-side to *p*-side of the junction, and holes in the reverse direction. Recombination of electron-hole pairs occurs with the emission of a photon, whose energy will be close to the band gap of the semiconductor.



materials (after Krames *et al.*, 2000; updated with UV LED shade rom interent 2002).



This process is not efficient for an indirect band-gap semiconductor such as Si or Ge, and so direct gap III - V or II - VI materials are commonly employed. Using materials with wider band gaps gives higher frequency light. In recent years, efficient LEDs have been developed across the visible spectrum (see Fig. 11.7, and are now more efficient than incandescent bulbs. Recent LED developments include not only wide-gap inorganic materials but also organic materials. These can be processed in different and simpler ways from inorganic compound semiconductors, have a larger intrinsic radiative coupling, and are of course flexible.

11.3 Field effect transistor

The ability of an external potential to shift the electronic energy levels is the basis of ubiquitous electronic technology of the field-effect transistor. Previously we discussed direct metal-semiconductor junctions, but if we place an insulating barrier between the metal and semiconductor, we can adjust the band-bending at the interface by amounts as large as the semiconductor gap itself, as shown in Fig. 11.8. The width of the well is often narrow enough that the levels within it are quantised.

Consequently a metal electrode placed over an oxide barrier on a Si substrate can be used as a gate to modulate the conductivity in the plane. This is the basis of the Metal-Oxide-Si-Field-Effect-Transistor, or MOSFET, the workhorse of modern electronics.



Figure 11.8: A Metal-Insulator-Semiconductor device allows a bias potential to be applied to bend the band edges of the semiconductor beside the insulating barrier. If the bending is large enough, a p-type semiconductor — as shown here — can have the conduction band pulled down below the chemical potential. This creates a narrow channel called an inversion layer, whose density can be modulated by the *gate* potential eV applied to a metal electrode.

11.4 Compound semiconductor heterostructures

Bandstructure engineering

Another way to make an inversion layer is to change the semiconductor chemistry in a discontinuous fashion within the same crystal structure. Epitaxial, atomic layer-by-layer growth allows the chemical composition and doping to be manipulated in fine detail. Such devices of compound semiconductors are used, for example, in semiconductor lasers for optical discs, in high speed electronics (e.g. cellphones) and high-speed lasers in telecommunications. This technology has also enabled fundamental science, by preparing very high mobility electron systems (e.g. for the quantum Hall effects), making "quantum wires" that are so thin as to have quantised levels, and for studies of the neutral electron-hole plasma as a possible superfluid.

Alloys of compound semiconductors, e.g. $Al_{1-x}Ga_xAs$, allow one to continuously vary the optical gap and the position of the band edges by varying the composition x. Two different semiconductors will - when referred to the vacuum potential at infinity - have bands that will in general line up in an offset fashion. We consider here only the case (like (Al, Ga)As when the band edges of one semiconductor lie entirely within the band gap of the other, though staggered overlaps do occur. When the materials are placed in contact, their Fermi energies must equalise, which is accomplished by charge transfer across the boundary. This lowers the conduction band edge on one side of the interface, and if doped sufficiently the band edge falls below the chemical potential, so that an equilibrium electron gas forms at the interface.

Inversion layers

Fig. 11.9 shows an outline of a scheme called modulation doping, where the donor levels are placed on the side of the interface away from the electron layers (and often at some distance from the interface). This has the advantage of creating an electron gas in a region where the crystal structure is nearly perfect, and mobilities greater than $10^3 m^2 V^{-1} s^{-1}$ have been achieved at low temperature. By addition of metal gates to the surface of the structures, electrical potential gradients can be applied to continuously vary the electron density in the layer, to pattern one dimensional wires, and to construct other interesting spatial structures.

Quantum wells

One of the most widespread applications of semiconductor multilayers is to make a quantum well — a thin region of a narrow gap material sandwiched inside a wide-gap one. Because the wells can be made very narrow, quantisation of the levels is important. In general, the eigenstates will be of the form $\Phi(\mathbf{r}, z) = \phi_n(z)e^{i\mathbf{k}\cdot\mathbf{r}}$ where \mathbf{r} and \mathbf{k} are here two-dimensional vectors, describing position and momentum in the plane. The situation for holes is more complex, because the degeneracy of the light and heavy hole states in bulk is broken by the 2D geometry. The details are important in practice, but not exciting.

Qu.11.34 Quantum well sub-bands

A 10 nm thick quantum well of GaAs is surrounded by bulk $Al_{0.7}Ga_{0.3}As$. The conduction band offset is 0.26 eV, and the effective mass of electrons in GaAs is 0.066 m_e .

(a) Estimate the energies of the (bottom of the) sub-bands $E_n(\mathbf{k} = 0)$, assuming the walls of the potential are infinitely high.

(b) What is the maximum areal density of electrons that can be occupied in the lowest sub-band before the second sub-band starts to be filled?

(c) How many sub-bands do you estimate exist for the actual sitation — a well of finite potential depth?

(d)* Note the word estimate in (c). Nevertheless, the 1D finite potential well is not a difficult problem to solve, though the actual solution of eigenstate energies needs to be done graphically.

For a potential of depth V_0 and width $L,\ {\rm show}\ {\rm that}\ {\rm the}\ {\rm number}$ of bound states is

$$1 + Int \left[(2m^* V_0 L^2 / \pi^2 \hbar^2)^{1/2} \right] \quad . \tag{11.7}$$



Figure 11.9: Formation of a 2D electron system by modulation doping. (a) Shows two semiconductors not in contact, with different chemical potentials, determined by the doping level. (b) is the band scheme that results when they are placed in contact. If the doping level is high enough - as shown here - the band edge on the left may fall below the chemical potential, and a layer of electrons is formed at the interface. (c) Shows the modulation doping scheme in more detail. Donors are placed to the right of the interface, so that the electron layer is pristine and free of impurities. The well width may be narrow enough that electron levels are quantised in a direction perpendicular to the barrier, forming sub-bands.

Quantum well laser

The operation of a laser requires an efficient mechanism for luminescent electronhole recombination, which rules out indirect gap semiconductors in practice. Lasing operation requires high densities of electrons and holes so that the probability of stimulated emission overcomes that of absorption. This latter condition requires *inversion*, meaning that the average electron (hole) occupancy in the luminescing states exceeds 1/2.

A double heterojunction laser is designed to achieve high densities, by using a quantum well — designed to trap both electrons and holes — with the source of carriers being a p-doped region on one side of the well, and an n-doped region on the other (see Fig. 11.10). This is indeed a diode (because holes can flow in from the p-side and electrons from the n, but not vice versa), but it is not operated in the same regime as a conventional diode. Instead, a rapid rate of recombination in the lasing region maintains different chemical potentials for electron and hole systems.



Figure 11.10: Operation of a double heterojunction laser. Notice the quasiequilibrium condition, with separate electron and hole chemical potentials.