

Willoughby Seago

Theoretical Physics

Introduction to Condensed Matter Physics

September 20, 2021

COURSE NOTES

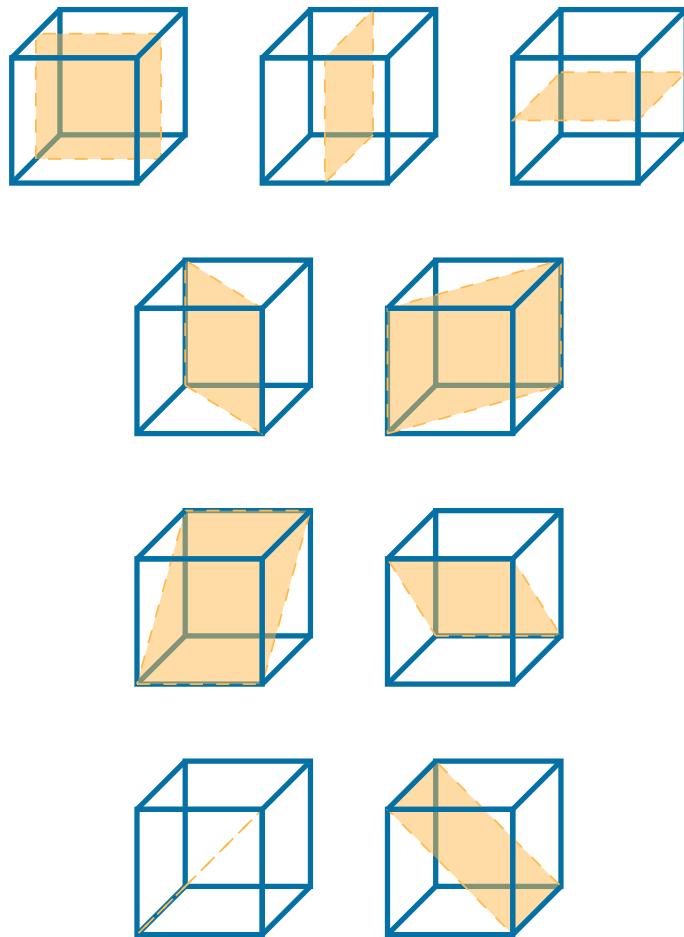
Introduction to Condensed Matter Physics

Willoughby Seago

September 20, 2021

These are my notes from the course introduction to condensed matter physics. I took this course as a part of the theoretical physics degree at the University of Edinburgh.

These notes were last updated at 11:50 on December 17, 2021. For notes on other topics see <https://github.com/WilloughbySeago/Uni-Notes>.



Chapters

	Page
Chapters	ii
Contents	iv
List of Figures	viii
List of Tables	x
1 Introduction	1
I Hard Condensed Matter	2
2 Crystals	3
3 Crystal Planes	13
4 Diffraction	16
5 Crystal Binding	28
6 Phonons	42
7 Free Electron Gas	65
8 Electron Energy Bands	82
9 Semiconductors	99
Appendices	113
A Constants and Useful Values	114
B Integrals	116
Bibliography	118

<i>CHAPTERS</i>	iii
Acronyms	119
Index	120

Contents

	Page
Chapters	ii
Contents	iv
List of Figures	viii
List of Tables	x
1 Introduction	1
I Hard Condensed Matter	2
2 Crystals	3
2.1 Crystal Definition	3
2.2 Periodic Crystals	3
2.2.1 Lattice	3
2.2.2 Symmetries	4
2.2.3 Symmetry Applications to Crystals	8
2.2.4 Basis	10
3 Crystal Planes	13
3.1 What Are They, and Why Do We Care?	13
3.2 Describing Planes	13
4 Diffraction	16
4.1 Experimental Setup	16
4.2 Diffraction Mechanism	16
4.2.1 Bragg's Law	16
4.2.2 The Maths of Diffraction	17
4.2.3 Fourier Series Recap	18
4.3 The Lattice Vector	19
4.4 Diffraction Intensities	20
4.5 Form Factors	22
4.6 Diffraction Techniques	23

4.6.1	Laue Diffraction	23
4.6.2	Monochromatic Single Crystal Diffraction	23
4.6.3	Monochromatic Powdered Crystal Diffraction	23
4.6.4	Atomic Form Factor	23
4.6.5	Neutron Diffraction	24
4.7	Systematic Absences	24
4.7.1	BCC Systematic Absences	24
4.8	FCC and BCC Reciprocal Lattices	25
4.9	Diffraction Beyond Simple Systems	25
4.9.1	DNA	25
4.9.2	Small Angle Scattering	26
4.9.3	Aperiodic Crystals	26
5	Crystal Binding	28
5.1	Van der Waals	29
5.1.1	Attractive Component	29
5.1.2	Repulsive Term	31
5.1.3	Lennard-Jones Potential	32
5.2	Van der Waals Crystal	33
5.2.1	Lattice Sums	34
5.2.2	Van der Waals Crystal Properties	35
5.3	Ionic Binding	36
5.3.1	Cohesive Energy	36
5.3.2	Ionic Crystals	36
5.3.3	Different Ionic Crystals	39
5.3.4	Ionic Spacing	39
5.4	Covalent Crystals	39
5.5	Metallic Crystals	41
5.6	Mixed Binding	41
6	Phonons	42
6.1	Heat Capacity	42
6.1.1	Classical Heat Capacity	42
6.1.2	Einstein Model	42
6.2	Phonons in One Dimension	43
6.3	Wigner–Seitz Cells	45
6.4	Phonons in Three Dimensions	45
6.5	Energy Quantisation and Momentum	48
6.6	Measuring Phonon Dispersion	48
6.7	Speed of Sound	49
6.8	Discrete Wave Vectors	49
6.8.1	Boundary Conditions in One Dimension	49
6.8.2	Boundary Conditions in Two Dimensions	50
6.8.3	Boundary Conditions in Three Dimensions	50
6.9	Phonon Density of States	50
6.9.1	Density of States in One Dimension	51
6.9.2	Density of States in Three Dimensions	52
6.10	Heat Capacity	52
6.10.1	Einstein Model	53
6.10.2	Debye Model	53

6.10.3 Beyond the Debye Model	56
6.11 Diatomic Linear Chain	56
6.11.1 Zone-Centre Solutions	57
6.11.2 Zone-Boundary Solutions	58
6.11.3 Acoustic and Optical Phonons	59
6.11.4 Polarisation	59
6.11.5 Multi-Atom Phonons in 3D	59
6.12 Thermal Expansion	59
6.13 Lattice Thermal Conductivity	61
6.13.1 Phonon Scattering	62
6.13.2 Temperature Dependence	63
7 Free Electron Gas	65
7.1 Electrons in a Box	66
7.1.1 One Dimension	66
7.1.2 Three Dimensions	67
7.2 Fermi Surface	68
7.3 Finite Temperatures	69
7.3.1 Density of States	70
7.4 Electronic Heat Capacity	71
7.4.1 Estimate	71
7.4.2 Electronic Heat Capacity	71
7.4.3 Thermal Effective Mass	74
7.5 Electrical Conductivity	75
7.5.1 Resistivity Temperature Dependence	76
7.6 Electronic Thermal Conductivity	78
7.7 Magnetic Field Effects	79
7.7.1 Hall Effect	79
8 Electron Energy Bands	82
8.1 Energy Bands	82
8.2 Band Gaps	82
8.3 Nearly Free Electron Model	83
8.4 Empty Lattice Approximation	85
8.5 Bloch's Theorem	88
8.5.1 Bloch's Theorem in One Dimension	90
8.6 Schrödinger Equation for a Periodic Potential	90
8.6.1 Bloch's Theorem Again	92
8.7 Fermi Surfaces	92
8.8 Higher Order Brillouin Zones	94
8.9 Fermi Surfaces in the First Brillouin Zone	94
8.10 Sharp Edges	98
9 Semiconductors	99
9.1 Band Structure	99
9.2 Semiconductors	100
9.3 Direct and Indirect Band Gaps	101
9.4 Wave Vectors	102
9.5 Holes	103
9.6 Effective Mass	105

9.6.1	In Three Dimensions	106
9.7	Typical Semiconductor Band Structures	106
9.7.1	Silicon	106
9.7.2	Germanium	106
9.7.3	Gallium Arsenide	106
9.8	Intrinsic Semiconductors	106
9.8.1	Conductivity	107
9.9	Temperature Dependence of the Fermi Level	109
9.10	Doping	110
9.11	Charge Carrier Density	110
9.12	Impurity as Hydrogen Like Centre	110
9.13	Semiconductor Devices	111
Appendices		113
A Constants and Useful Values		114
B Integrals		116
B.1	$\int_0^{\infty} x^{p-1} / (e^x - 1) dx$	116
Bibliography		118
Acronyms		119
Index		120

List of Figures

	Page
2.1 Possible square lattice vectors in 2D	4
2.2 Unit cell and lattice vectors.	5
2.3 Mirror plane symmetries of a cube.	6
2.4 Five fold rotational.	6
2.5 Rotational symmetry of a cube	7
2.6 Square, rectangular, and hexagonal lattice vectors in 2D	8
2.7 Primitive and standard centred rectangular lattice.	9
2.8 BCC crystal	11
2.9 FCC crystal.	12
3.1 Crystal planes examples.	14
4.1 Diffraction mechanism	17
4.2 Aperiodic Penrose tiling.	27
5.1 Dipole oscillators and van der Waals interactions.	29
5.2 Lattice distances for lattice sums.	34
5.3 One-dimensional ionic crystal.	38
5.4 Common ionic crystal structures.	40
5.5 Molecular orbital diagrams for H_2 and He_2	41
6.1 Heat capacity predicted by Einstein model.	44
6.2 One-dimensional chain of atoms.	44
6.3 Wigner–Seitz cell.	46
6.4 Wigner–Seitz cell tiles the plane	46
6.5 3D Brillouin zone	47
6.6 Dispersion relation	47
6.7 Diatomic linear chain dispersion relation.	57
6.8 Umklapp scattering.	63
7.1 Ionic potentials approximate an infinite square well	66

7.2 In blue is $D(\varepsilon)f(\varepsilon)$ at $T > 0$, in yellow is $D(\varepsilon)f(\varepsilon)$ at $T = 0$, the black line is $D(\varepsilon)$, and the dashed line is $f(\varepsilon)$. When increasing the temperature from 0 to some finite temperature we can do so in two steps, first taking all electrons that will become excited to energy ε_F , and then to their final, higher energy.	72
8.1 Band structure for different materials	83
8.2 Energy bands separated by a band gap.	84
8.3 ε_k and all shifts of ε_k by some lattice vector G	86
8.4 ε_k and all G -shifts of ε_k for some reciprocal lattice vector G	86
8.5 Reducing dispersion relations to the first Brillouin zone.	87
8.6 Electron bands in two dimensions.	88
8.7 Paraboloids intersect plane to give parabolas.	89
8.8 Square unit cell Fermi surfaces	93
8.9 Reduced dispersion relations.	95
8.10 Higher order Brillouin zones	96
8.11 Brillouin zones have the same area	97
8.12 Reduced Fermi surface.	97
8.13 Fermi surface coloured by Brillouin zone.	97
9.1 Model band structure.	99
9.2 Metal and insulator band structure.	100
9.3 Semimetal band structure.	101
9.4 Direct and indirect band gaps.	102
9.5 Absorbance of direct and indirect band gaps.	103
9.6 Hole band.	104
9.7 A p - n junction.	112
A.1 Periodic table.	115

List of Tables

	Page
2.1 2D Bravais lattices.	9
2.2 3D Bravais lattices.	9
2.3 Crystal structure centring.	10
8.1 The value of ε at $\mathbf{k} = \mathbf{0}$ and $\mathbf{k} = k_x \mathbf{a}_1^*$ for all G -shifts with $ G \leq \sqrt{2}(2\pi/a)$. Taking $\hbar^2/(2m) = 1$ for simplicity.	89
9.1 Summary of hole properties.	105
9.2 Typical band gaps and electron concentrations.	109
A.1 Constants.	114
A.2 Unit conversions.	114

One

Introduction

Condensed matter physics is the study of materials made of a large number of particles with strong interactions (as opposed to, say a gas, where particles interactions are weak). Both solids and liquids are considered to be condensed matter, with solids being hard condensed matter and liquids soft condensed matter. As well as this condensed matter also considers gels, suspensions, and many other forms of matter.

In this course we mostly focus on hard condensed matter, in particular on crystals as one of the simplest forms of condensed matter. The symmetry of even the most irregular crystals makes them much simpler than amorphous solids, or liquids.

As a part of hard condensed matter we will consider: crystal structures, diffraction, crystal binding, lattice vibrations, electron gases, electron energy bands, and semiconductors. For soft condensed matter we will consider: diffusion, and interactions in soft matter.

Condensed matter is unlike many other areas of physics. For example, in electromagnetism once you have Maxwell's equations all of electromagnetism follows, similarly with quantum mechanics everything stems from Schrödinger's equation, and in general relativity everything can be derived from Einstein's field equations. This isn't the case in condensed matter since we consider systems of many, say 10^{20} , particles and we consider the quantum mechanical interaction of these particles. It's hard enough to do this for a single atom of a simple element, like helium. Doing so for this many particles is completely infeasible. For this reason we focus on methods for modelling crystals and the results that we can achieve rather than the mathematical minutiae of each model.

Part I

Hard Condensed Matter

Two

Crystals

2.1 Crystal Definition

Crystals were first discovered as large collections of minerals with very specific angles and facets. As far back as the 16th century it was suggested that this large scale order was due to order on a smaller scale. We now know that crystals occur due to ordered arrangements of atoms. It was shown in the 18th century that it was possible to combine small cubes in an ordered way to build up the shapes seen on the macroscopic scale. It was confirmed via x-ray diffraction in the 20th century that crystals consist of an ordered, repeating arrangement of atoms.

Definition 2.1.1 — Crystal A **crystal** is a solid with an essentially ordered arrangement of atoms or molecules extending in all three spatial directions with an x-ray diffraction pattern consisting of essentially sharp dots.

Note that the use of the word “essentially” here is just to allow for slight defects in real crystals. We will mostly consider ideal crystals free from such defects. In fact, we will mostly restrict our study to a subset of crystals known as periodic crystals.

Definition 2.1.2 — Periodic Crystal A **periodic crystal** is a solid with an essentially ordered arrangement of atoms extending periodically in space.

2.2 Periodic Crystals

A periodic crystal can be described fully with two concepts. A lattice, which is a mathematical abstraction of its long range symmetry, and a basis, which determines how atoms are positioned with respect to the lattice.

2.2.1 Lattice

A **crystal lattice** is a regular array of points in space. They describe the three-dimensional crystal periodicity. It is an abstract mathematical set of points, there are no atoms involved in the definition of the lattice.

We can define a lattice through a set of **lattice vectors** which span the lattice. If we denote these lattice vectors \mathbf{a}_i then the **lattice points** are all points of the form

$$n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3 \quad (2.2.1)$$

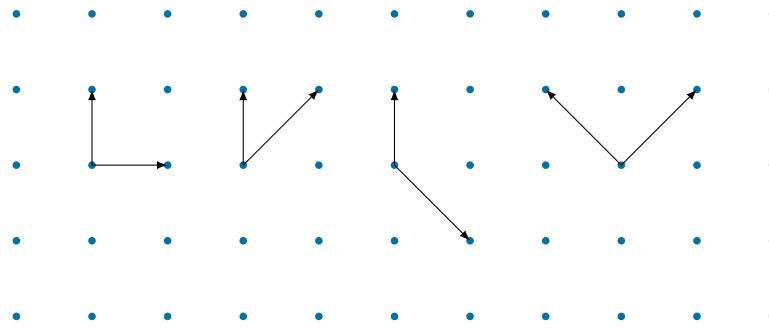


Figure 2.1: Possible pairs of lattice vectors defining the lattice points shown. Typically we wish to choose lattice vectors to be as short as possible and, if possible, orthogonal. This makes the first pair on the left the obvious choice.

where $n_i \in \mathbb{Z}$. This is in three dimensions, we often consider two dimensional crystals as they are easier to visualise, in this case we simply drop the third term and have only two lattice vectors. Note that this assumes the lattice extends to infinity in all directions. Obviously real crystals don't do this but they are often large enough that this is a reasonable approximation.

A crucial feature of crystals is that if we shift the crystal by

$$\Delta R = m_1 \mathbf{a}_1 + m_2 \mathbf{a}_2 + m_3 \mathbf{a}_3 \quad (2.2.2)$$

for $m_i \in \mathbb{Z}$ then the set of lattice points is unchanged. This translational symmetry is lacked by other forms of condensed matter, such as amorphous solids.

Note that the choice of lattice vectors is not unique. Therefore it is best to pick lattice vectors that simplify our analysis. Typically this means orthogonal vectors as short as possible. Some examples for a two-dimensional lattice are shown in [Figure 2.1](#)

We call the parallelepiped (or in 2D parallelogram) spanned by \mathbf{a}_1 , \mathbf{a}_2 , and \mathbf{a}_3 the **unit cell**. It has lattice points at its corners, and possibly inside and on its edges/faces. It is important to note that the points that are on the boundary are shared by neighbouring unit cells and therefore we must be careful not to over count. We call a unit cell with a single lattice point a **primitive unit cell**. For example considering the lattice vectors in [Figure 2.1](#) the first three define primitive unit cells whereas the second defines a unit cell with 2 lattice points, 4 in the corners, each shared with 3 other neighbours, and one in the centre.

A unit cell is characterised by the lengths of the lattice vectors and the angles in between. That is, the unit cell is characterised by $a = |\mathbf{a}_1|$, $b = |\mathbf{a}_2|$, $c = |\mathbf{a}_3|$, and $\alpha = \angle(\mathbf{a}_2, \mathbf{a}_3)$, $\beta = \angle(\mathbf{a}_1, \mathbf{a}_3)$, and $\gamma = \angle(\mathbf{a}_1, \mathbf{a}_2)$, where $\angle(\mathbf{p}, \mathbf{q})$ is the angle between \mathbf{p} and \mathbf{q} . See [Figure 2.2](#). This is the standard notation for these quantities and we will use it from here on without defining them.

The volume of the unit cell is given by the scalar triple product:

$$V = |\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)| \quad (2.2.3)$$

2.2.2 Symmetries

We can learn a lot about a crystal from the symmetries that it possess, or fails to posses. We can see the symmetry of many crystals from the shapes that they form.

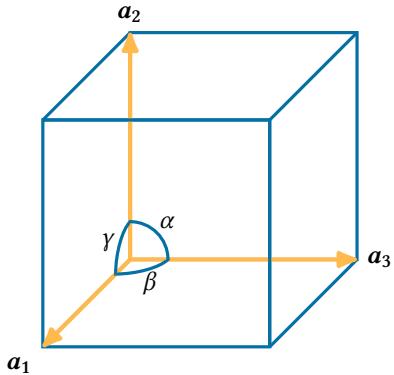
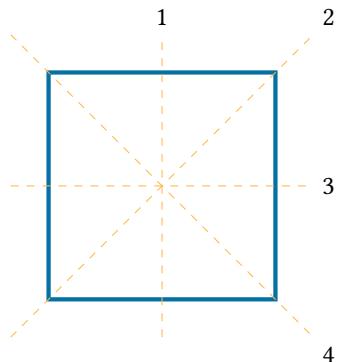


Figure 2.2: A unit cell spanned by the lattice vectors \mathbf{a}_1 , \mathbf{a}_2 , and \mathbf{a}_3 .

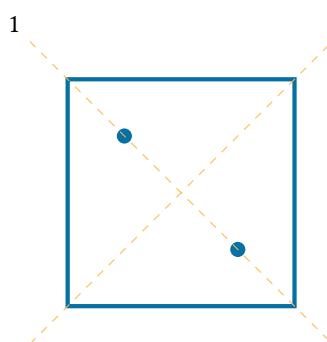
For example, cubic pyrite, FeS_2 , forms cubic crystals. These have four-fold rotational symmetry about their faces, and also three fold rotational symmetry about their points. There are a few types of symmetry that are interesting for crystals and we shall discuss these in this section.

2.2.2.1 Mirror Plane Symmetry

A **mirror plane symmetry** means that there is a plane (or in two dimensions a line) through which we can reflect the object and the result is identical to having done nothing. For example a simple square has four mirror symmetries:



If we introduce another feature, such as dots, we reduce the number of symmetries:



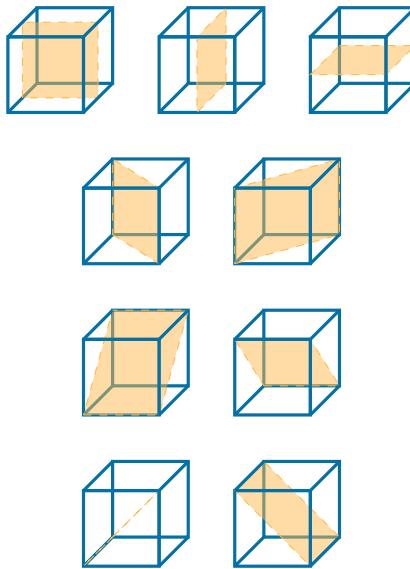


Figure 2.3: Mirror plane symmetries of a cube, there are 9 in total.



Figure 2.4: A five pointed star has 5-fold rotational symmetry, including a rotation by 0 or 2π , which is the identity, in that it leaves the shape completely the same.

A cube similarly has mirror plane symmetries, in fact, it has 9 of them. This is shown in [Figure 2.3](#).

2.2.2.2 Rotational Symmetry

An object has k -fold **rotational symmetry** about some axis, \hat{n} if it is possible to rotate the object by some multiple of $2\pi/k$ about \hat{n} and it looks like nothing has happened. For example, [Figure 2.4](#) shows a five pointed star with 5-fold rotational symmetry, including a rotation by 0 or 2π , which is equivalent to doing nothing.¹

A square has four-fold rotational symmetry.² A cube also has rotational symmetries, but in three dimensions things are more complicated as we have far more choices of axes of rotation. [Figure 2.5](#) shows a cube with three axes of rotational symmetry. The first connects two diagonally opposite corners. The cube has three-fold rotational symmetry about this axis. The second connects the mid points of two

¹The symmetry group is \mathbb{Z}_5 , or more generally for a k -fold rotationally symmetric object, \mathbb{Z}_k .

²The dihedral group, D_4 , is the symmetry group of a square combining the rotational and reflective symmetries.

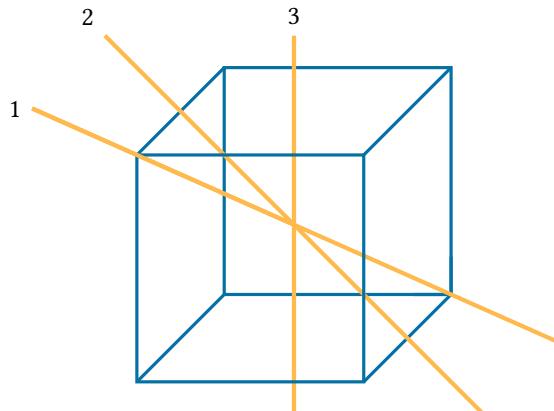
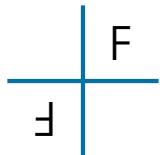


Figure 2.5: Three possible axes of rotational symmetry of the cube.

diagonally opposite edges. The cube has two-fold rotational symmetry about this axis. The third connects the centre of two opposite faces. The cube has four-fold rotational symmetry about this axis. There are four axes of the first type, six of the second type, and three of the third type, for 13 total axes of rotation.

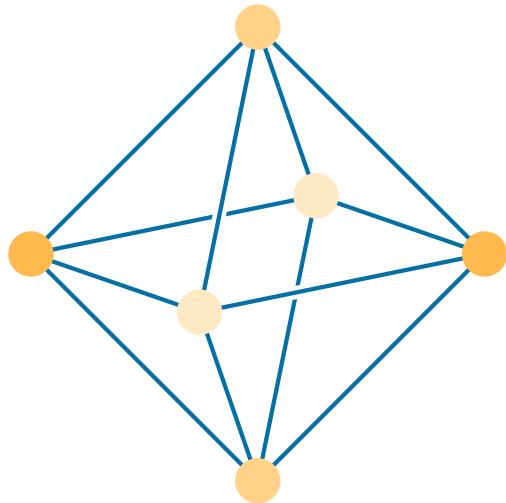
2.2.2.3 Inversion

An **inversion symmetry** is reflection through a point, called the **centre of inversion**, which leaves the object unchanged. For example, this graph,



has a centre of inversion at the origin.

This tetrahedron,



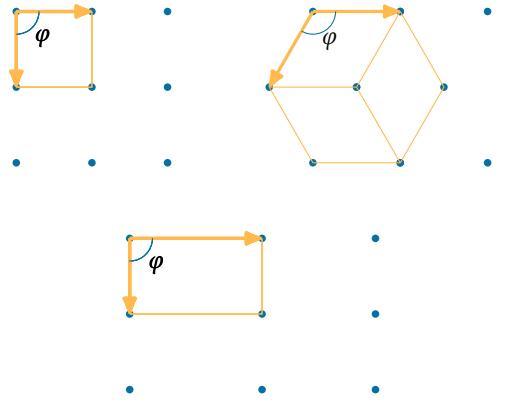


Figure 2.6: A square lattice (top left), hexagonal lattice (top right), and hexagonal lattice (bottom). For the square and rectangular lattices $\varphi = \pi/2$ and for the hexagonal lattice $\varphi = 2\pi/3$. For the square and hexagonal lattice $|\mathbf{a}_1| = |\mathbf{a}_2|$.

has a centre of inversion at its centre. The corresponding symmetry swaps points of the same shade.

A cube has a centre of inversion at its centre. Combining this with the 9 mirror plane symmetries and 13 rotational symmetries we see a cube has 24 symmetries to consider.

2.2.3 Symmetry Applications to Crystals

Suppose a crystal has four-fold rotational symmetry. It is not difficult to convince yourself that this can only be the case if $|\mathbf{a}_1| = |\mathbf{a}_2|$ and $\angle(\mathbf{a}_1, \mathbf{a}_2) = \pi/2$. We call this a square lattice for obvious reasons.

If we only require two-fold rotational symmetry then we can relax the condition on the vector lengths and we get a rectangular lattice, which has $|\mathbf{a}_1| \neq |\mathbf{a}_2|$ and $\angle(\mathbf{a}_1, \mathbf{a}_2) = \pi/2$.

The symmetry imposes conditions on the lattice parameters. Similarly if a crystal has three-fold rotational symmetry then this necessitates that $|\mathbf{a}_1| = |\mathbf{a}_2|$ and $\angle(\mathbf{a}_1, \mathbf{a}_2) = 2\pi/3$. We call this a hexagonal lattice as three unit cells together form a regular hexagon.

These three lattices, along with relevant unit cells, are shown in [Figure 2.6](#), as well as the three unit cells forming a hexagon in the hexagonal lattice.

What we are discussing are two-dimensional lattices. The three we have discussed are the result of symmetry restrictions. There is also a lattice that results from no symmetry restrictions, called the centred rectangular lattice. This is shown in [Figure 2.7](#). The primitive unit cell shown is awkward to work with and doesn't lead to much insight. If we allow ourselves to work with non-primitive unit cells then the second unit cell shown is a more sensible choice. This is called a centred rectangular unit cell, since it is rectangular and has a lattice point at its centre.

The four two-dimensional lattices we have seen cover all possible cases. These are together known as the two-dimensional **Bravais lattices**. They are summarised in [Table 2.1](#).

In three dimensions we can make similar arguments about symmetries and the resulting restrictions on the lattice parameters. We end up with a list of seven three-dimensional Bravais lattices. The key information for these is in [Table 2.2](#).

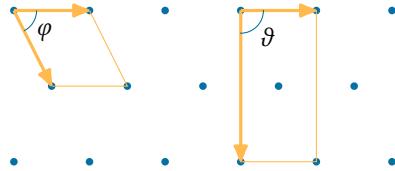


Figure 2.7: The centred rectangular lattice with its primitive unit cell and the simpler choice of a centred rectangular unit cell.

Table 2.1: The key information about two-dimensional Bravais lattices.

Crystal	External Minimum Symmetry	Restrictions
Square	Four-fold rotational symmetry	$ \mathbf{a}_1 = \mathbf{a}_2 $ $\angle(\mathbf{a}_1, \mathbf{a}_2) = \pi/2$
Rectangular	Two-fold rotational symmetry	$\angle(\mathbf{a}_1, \mathbf{a}_2) = \pi/2$
Hexagonal	Three-fold rotational symmetry	$ \mathbf{a}_1 = \mathbf{a}_2 $ $\angle(\mathbf{a}_1, \mathbf{a}_2) = 2\pi/3$
Centred Rectangular	None	None

Table 2.2: The key information about three-dimensional Bravais lattices.

Crystal	External Minimum Symmetry	Restrictions
Cubic	Four three-fold rotational symmetries	$a = b = c$ $\alpha = \beta = \gamma = \pi/2$
Hexagonal	One six-fold rotational symmetry	$a = b, \gamma = 2\pi/3$ $\alpha = \beta = \pi/2$
Trigonal	One three-fold rotational symmetry	$a = b = c$ $\alpha = \beta = \gamma \leq 2\pi/3$ $\alpha \neq \pi/2$
Tetragonal	One four-fold rotational symmetry	$a = b$ $\alpha = \beta = \gamma = \pi/2$
Orthorhombic	Three perpendicular two-fold rotational symmetries	$\alpha = \beta = \gamma = \pi/2$
Monoclinic	One two-fold rotational symmetry, parallel to b	$\alpha = \gamma = \pi/2 \neq \beta$
Triclinic	None	None

Table 2.3: Crystal systems and centring. Note that R stands for rhombohedral, a trigonal crystal can be categorised as having either a hexagonal or rhombohedral lattice system

Crystal System	Centring
Cubic	P, F, and I
Hexagonal	P
Trigonal	P, and R
Tetragonal	P, and I
Orthorhombic	P, I, and F
Monoclinic	P, and A/B/C
Triclinic	P

This isn't quite the complete picture however as we have assumed these are all primitive unit cells. In three dimensions there are actually four centring, where we add additional lattice points to the unit cell, to consider. The first, and simplest, is the primitive unit cell, P. The second is face centring, F, where we add an extra lattice point to every face. The third is body centring, I, where we add an extra lattice point to the centre of the cell (I for interior). The fourth is base centred, denoted either A, B, or C, where we add an extra lattice point to one pair of opposite faces. If we add to the face spanned by \mathbf{a}_1 and \mathbf{a}_2 then we call it C, the face spanned by \mathbf{a}_1 and \mathbf{a}_3 corresponds to B, and the face spanned by \mathbf{a}_2 and \mathbf{a}_3 corresponds to A. Not all crystal system and centring pairings give us a useful lattice however. The ones that do are given in [Table 2.3](#). We are now in a position to fully specify a crystal lattice by stating the crystal structure and the centring.

2.2.4 Basis

Now that we can fully specify a crystal lattice we need to use this to define a crystal. We do this with the help of a basis, which is a list of atoms and locations that they occur in the unit cell. We typically give these locations in dimensionless, reduced coordinates, (x, y, z) , corresponding to the position

$$\mathbf{r} = x\mathbf{a}_1 + y\mathbf{a}_2 + z\mathbf{a}_3 \quad (2.2.4)$$

The simplest basis is to have only a single atom at $\mathbf{r} = (0, 0, 0)$. This is the case with many pure metals, for example Na, Ba, and Fe all have body-centred cubic lattices (BCC) with this basis, and Ca, Ni, and Au have face-centred cubic lattices (FCC) with this basis. It is worth looking at these examples in more detail.

2.2.4.1 Body-Centred Cubic

A body-centred cubic lattice unit cell consists of lattice points at the eight corners of a cube, and one lattice point in the middle. We consider the case of the basis being a single atom at $(0, 0, 0)$. The conventional lattice vectors in this case are $\mathbf{a}_1 = a\hat{x}$, $\mathbf{a}_2 = a\hat{y}$, and $\mathbf{a}_3 = a\hat{z}$, where \hat{x} , \hat{y} , and \hat{z} are orthonormal Cartesian unit vectors. There are two atoms in the conventional unit cell, one at the centre, and $1/8$ of an atom in each corner, these are shared between the neighbouring cells. The positions of the atoms in reduced coordinates in the unit cell are $(0, 0, 0)$ and $(1/2, 1/2, 1/2)$. Each atom has 8 nearest neighbours and 6 next-nearest neighbours.

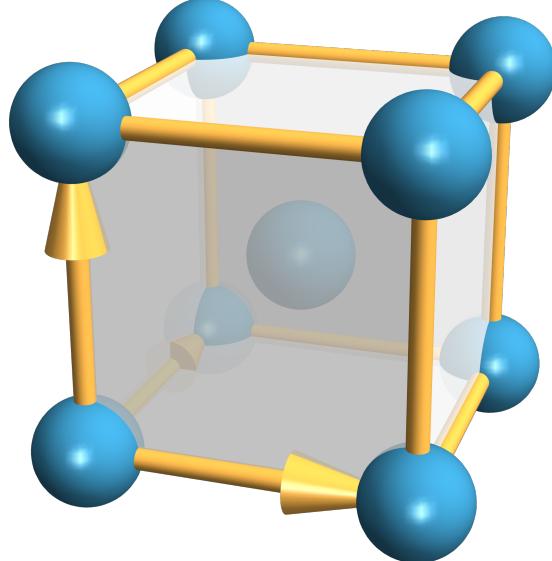


Figure 2.8: A body-centred crystal.

The primitive unit cell is more complicated so we don't use it often but it has lattice vectors

$$\mathbf{a}_1^P = \frac{a}{2}(-\hat{x} + \hat{y} + \hat{z}), \quad \mathbf{a}_2^P = \frac{a}{2}(\hat{x} - \hat{y} + \hat{z}), \quad \text{and} \quad \mathbf{a}_3^P = \frac{a}{2}(\hat{x} + \hat{y} - \hat{z}). \quad (2.2.5)$$

It is clear why one would prefer a centred cubic lattice over this oddly shaped primitive lattice.

2.2.4.2 Face-Centred Cubic

A face-centred cubic lattice unit cell consists of a lattice point at the eight corners of a cube, and one lattice point in the centre of each face. We consider the case of a basis of a single atom at $(0, 0, 0)$. The conventional lattice vectors in this case are $\mathbf{a}_1 = a\hat{x}$, $\mathbf{a}_2 = a\hat{y}$, and $\mathbf{a}_3 = a\hat{z}$, where \hat{x} , \hat{y} , and \hat{z} are orthonormal Cartesian unit vectors. There are four atoms in the conventional unit cell. Each of the six faces contributes half an atom, and the four corners contribute $1/8$ of an atom. The positions of the atoms in reduced coordinates are $(0, 0, 0)$ and $(1/2, 0, 0)$. Each atom has 12 nearest neighbours.

The primitive unit cell is more complicated so we don't use it often but it has lattice vectors

$$\mathbf{a}_1^P = \frac{a}{2}(\hat{y} + \hat{z}), \quad \mathbf{a}_2^P = \frac{a}{2}(\hat{x} + \hat{z}), \quad \text{and} \quad \mathbf{a}_3^P = \frac{a}{2}(\hat{x} + \hat{y}). \quad (2.2.6)$$

It is clear why one would prefer a centred cubic lattice over this oddly shaped primitive lattice.

The face-centred cubic lattice is the densest possible way to pack hard, non-overlapping spheres. For this reason it is also called **cubic close packed** (CCP).

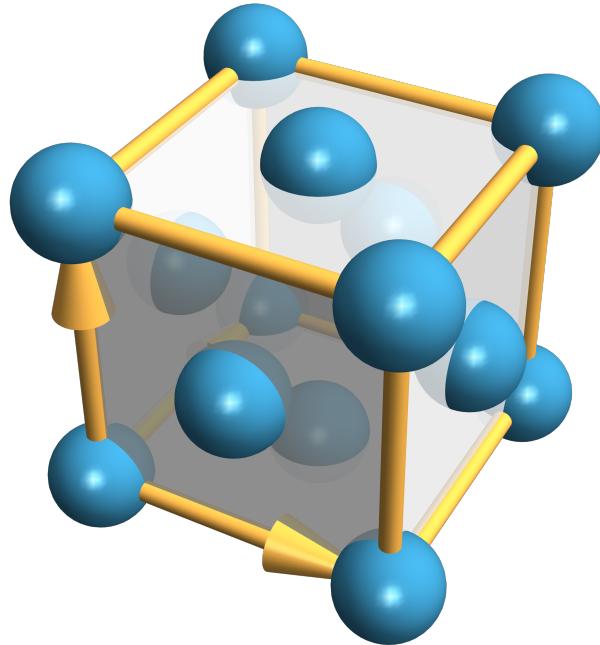


Figure 2.9: A face-centred crystal.

2.2.4.3 Hexagonal Close-Packed

A hexagonal close-packed (HCP) consists of a hexagonal unit cell, with lattice parameters $a = b$ and $\gamma = 2\pi/3$. There are two atoms in the conventional unit cell, positioned at $(2/3, 1/3, 1/4)$ and $(1/3, 2/3, 3/4)$. Each atom has 12 nearest neighbours. The packing density of this structure is equal to that of cubic close-packing. There is no conventional choice of lattice vectors, but one option is

$$\mathbf{a}_1 = a \frac{\sqrt{3}}{2} \hat{x} + \frac{a}{2} \hat{y}, \quad \mathbf{a}_2 = -a \frac{\sqrt{3}}{2} \hat{x} + \frac{a}{2} \hat{y}, \quad \text{and} \quad \mathbf{a}_3 = c \hat{z}. \quad (2.2.7)$$

2.2.4.4 Close Packing

Both cubic and hexagonal close packing can be created by layering hexagonally arranged layers of spheres. The difference comes in where you choose to place the third layer. If you choose to place the third layer such that it lines up with the first layer then the result will be hexagonal close packing. If you choose instead to place it so that it doesn't line up you will get cubic close-packing.

2.2.4.5 Real Life Example

Consider NaCl, common table salt, this has a face-centred cubic lattice. We need to choose an atom type to have at the position $(0, 0, 0)$, we arbitrarily choose Cl. The basis is then a Cl atom at $(0, 0, 0)$ and a Na atom at $(1/2, 0, 0)$, or equivalently at $(1/2, 1/2, 1/2)$.

Crystals can become very complicated in the real world but they can always be broken down into a choice of lattice and centring and a basis of atoms.

Three

Crystal Planes

3.1 What Are They, and Why Do We Care?

Crystal planes are sets of parallel planes upon which all of the atoms in a crystal can be located. The nature of the planes in a crystal can be important for its physical properties. The most obvious example being diffraction, and in particular Bragg's law, which predicts from the spacing between parallel planes what diffraction pattern will occur. We will see this later when we consider diffraction in more detail.

Transport properties (such as conductivity, both thermal and electrical) are also dependent on planes. Some materials have anisotropic transport properties, that is they may be better at transport in one direction. For example, graphite is made of layers of planes in which the atoms are arranged hexagonally. Graphite is a much better as an electrical and thermal conductor along the planes rather than across the planes.

Planes can also be important for the structural properties of materials. A crystal is much more likely to break along a plane and plastic deformation, such as a metal being permanently bent, most often occurs when planes slip past each other.

The crystal planes also effect the surface properties of a crystal. For example, they may effect the ability of the crystal to act as a catalyst or semiconductor.

3.2 Describing Planes

Miller indices provide a standard method of naming planes. The naming scheme goes as follows. First, find where the plane intersects each axis in reduced units, say, at $(3, 2, 2)$. If there is no intercept then we say the intercept occurs at ∞ , and we take $1/\infty = 0$. Second, take the reciprocals, so for our example, $(1/3, 1/2, 1/2)$. Third, multiply through by the smallest integer that leaves all whole numbers, in our case this is six and we get $(2\bar{3}3)$. This is the name of our plane. If there is a negative number then we denote this with a bar above the number.

The crystal plane $(h k \ell)$ intercepts the real space axes at a_1/h , a_2/k , and a_3/ℓ .

The Miller indices don't represent a single plane but an infinite family of parallel planes.

We can also use Miller indices to denote a direction in a crystal. We denote by $[h k \ell]$ the direction $ha_1 + ka_2 + \ell a_3$. In cubic, tetragonal, and orthorhombic lattices, where all angles are $\pi/2$, the direction $[h k \ell]$ is simply the direction perpendicular to the plane $(h k \ell)$.

Consider the planes shown in [Figure 3.1](#). In particular $(1\ 0\ 0)$, $(2\ 0\ 0)$, and $(\bar{1}\ 0\ 0)$. Notice that $(1\ 0\ 0)$ and $(\bar{1}\ 0\ 0)$ describe the same set of planes. That is planes perpen-

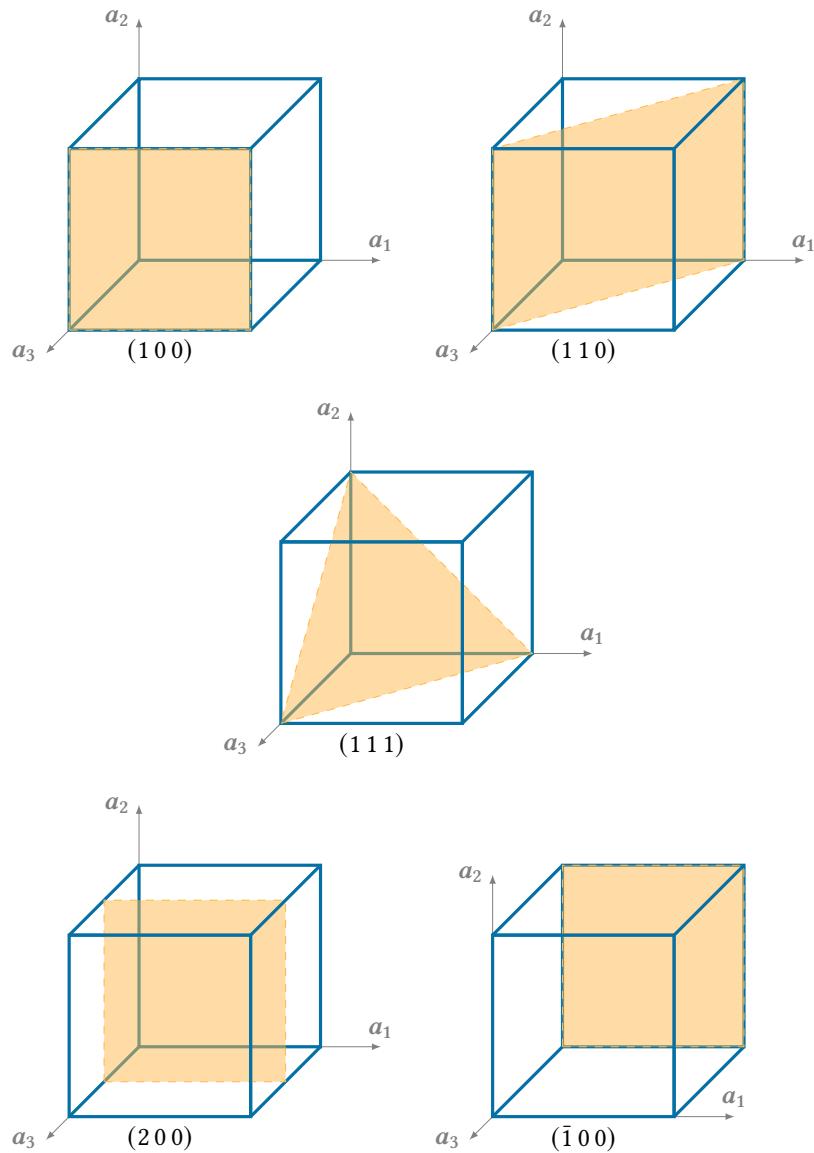


Figure 3.1: Some crystal planes labelled by their Miller indices. Note that $(1\ 0\ 0)$, $(2\ 0\ 0)$, and $(\bar{1}\ 0\ 0)$ are all parallel. The family of planes described by $(1\ 0\ 0)$ and $(\bar{1}\ 0\ 0)$ are the same and $(2\ 0\ 0)$ describes these planes but also the planes directly between.

dicular to α_1 and spaced a distance a apart. This is because these examples are for a cubic lattice and the high symmetry means lots of planes have multiple descriptions.

The planes $(2\ 0\ 0)$ are also parallel to α_1 but they are spaced $a/2$ apart. This means that the family of planes $(2\ 0\ 0)$ contains all the planes described by $(1\ 0\ 0)$, and also all of the planes halfway between.

Four

Diffraction

4.1 Experimental Setup

Crystal **diffraction** is most often done with x-rays, although it can be done with neutrons or electrons. The wavelength needs to be on the same order as the interatomic spacing, $\sim 1 \text{ \AA}$. Although, electrons tend to interact too strongly to be useful for larger samples as they have too many secondary interactions. Electrons and x-rays interact with the electron shells of the atoms whereas neutrons interact with the nuclei, although this makes little difference in the final outcome.

The general setup is to have a small amount of the crystal placed in the beam of monochromatic x-rays. Behind the sample is a screen that can detect x-rays. The crystal is rotated in the beam and at different rotations the x-rays are diffracted to different points on the screen. In doing this a picture is built up of all possible diffraction angles. We can also measure the intensity at each diffraction angle.

In general we will end up with a pattern of spots on the screen. The intensities of the spots will not all be the same. It is possible to work backwards from the spots and tell what planes were involved in the diffraction. Some planes aren't ever involved in the diffraction. These are called systematic absences.

4.2 Diffraction Mechanism

Diffraction is a wave phenomenon. Incident radiation is scattered in all directions by the atoms. The order of a crystal leads to interference, both constructive and destructive, which results in only some directions

4.2.1 Bragg's Law

Bragg's law states that for interplanar spacing d_{hkl} and radiation wavelength λ there will be constructive interference at angle 2θ if

$$2d_{hkl} \sin \theta = \lambda \quad (4.2.1)$$

This is the crystallographer's version of Bragg's law, normally there is a factor of $n \in \mathbb{Z}$ but this isn't needed when we consider whole families of planes at a time.

Bragg's derivation of this law assumes that the crystal planes act like perfect mirrors. Further, this law gives us no information about the diffracted intensities and cannot account for systematic absences.

Bragg's law does allow us to determine the unit cell dimensions and hence the lattice part of the crystal structure. What it can't do is tell us anything about the

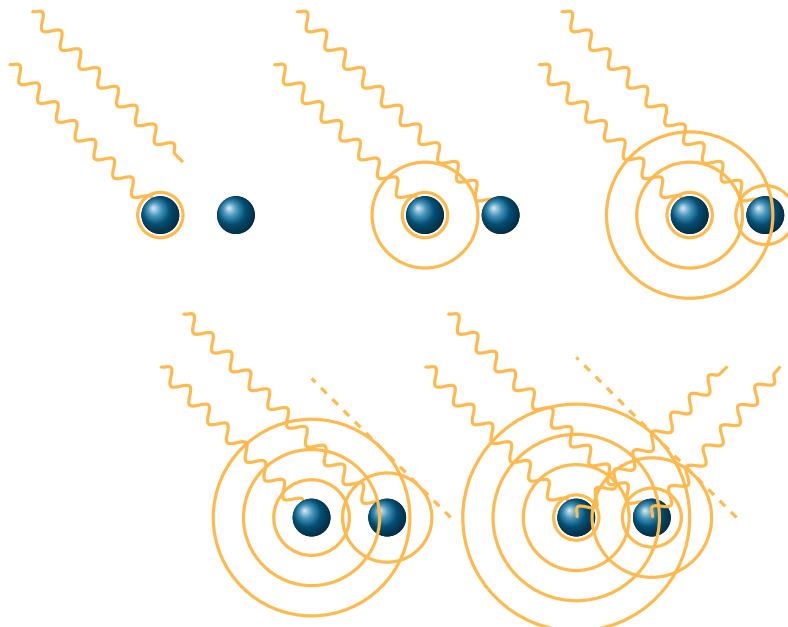


Figure 4.1: Two atoms in a crystal absorb and re-emit light.

arrangement of atoms in this lattice, i.e. the crystal basis. For this we will need a fuller description of the diffraction process.

4.2.2 The Maths of Diffraction

Suppose the incident light is an electromagnetic plane wave with wave vector \mathbf{k} , that is proportional to $e^{i\mathbf{k}\cdot\mathbf{r}}$. Similarly the diffracted beam is an electromagnetic plane wave with wave vector \mathbf{k}' , and hence proportional to $e^{i\mathbf{k}'\cdot\mathbf{r}}$.

We now note that $|\mathbf{k}| = 2\pi/\lambda$, where λ is the wavelength of the radiation. The phase of the wave at \mathbf{r} is given by $\mathbf{k} \cdot \mathbf{r}$. In a diffraction process there is no change in wavelength and so $|\mathbf{k}| = |\mathbf{k}'| = k$. This assumes elastic scattering, meaning the total energy is constant, and since $E = \hbar ck$ this means k is constant. This is a reasonable assumption as any effects due to inelastic scattering are necessarily weaker due to loss of energy.

Assuming x-ray diffraction the diffraction occurs due to interaction with the electrons. We can describe the positions of the electrons with the electron density, $n(\mathbf{r})$.

We consider the phase shift at each point in the crystal. The phase shift at some point \mathbf{r} is $(\mathbf{k} - \mathbf{k}') \cdot \mathbf{r}$. We can account for the phase shift from each point of the crystal by integrating over the volume of the crystal. This gives us the **scattering amplitude**:

$$F(\mathbf{k}, \mathbf{k}') = \int_{\text{crystal}} n(\mathbf{r}) \exp[i(\mathbf{k} - \mathbf{k}') \cdot \mathbf{r}] dV. \quad (4.2.2)$$

This measures the strength of diffraction of a wave entering in the direction $\hat{\mathbf{k}}$ and exiting in the direction $\hat{\mathbf{k}'}$.

For periodic crystals $n(\mathbf{r})$ is periodic in space. This suggests that we should use Fourier series.

4.2.3 Fourier Series Recap

4.2.3.1 One Dimension

A function is periodic if $n(x+a) = n(x)$ for all x for some specific value of a . We call a the period of the function. A periodic function can be expanded as a Fourier series. In the one-dimensional case the function n can be expanded as

$$n(x) = \frac{1}{2}C_0 + \sum_{p=1}^{\infty} \left[C_p \cos\left(\frac{1\pi px}{a}\right) + S_p \sin\left(\frac{2\pi px}{a}\right) \right]. \quad (4.2.3)$$

The coefficients are given by

$$C_p = \frac{2}{a} \int_0^a n(x) \cos\left(\frac{2\pi px}{a}\right) dx, \quad (4.2.4)$$

$$S_p = \frac{2}{a} \int_0^a n(x) \sin\left(\frac{2\pi px}{a}\right) dx. \quad (4.2.5)$$

We can similarly expand n in a complex Fourier series:

$$n(x) = \sum_{p=-\infty}^{\infty} n_p e^{2\pi ipx/a} \quad (4.2.6)$$

where

$$n_p = \frac{1}{a} \int_0^a n(x) e^{-2\pi ipx/a} dx. \quad (4.2.7)$$

If n is a real function then $n_{-p} = n_p^*$.

Notice that if x is a length then p has units of inverse length.

4.2.3.2 Higher Dimensions

In two or more dimensions we can do something similar. Consider a two-dimensional function n which is periodic in both of its inputs. Then the Fourier series of this function follows very easily from the one dimensional case:

$$n(x, y) = \sum_{p,q=-\infty}^{\infty} n_{p,q} e^{2\pi ipx/a} e^{2\pi iqy/b} \quad (4.2.8)$$

where a is the period along the x -axis and b is the period along the y -axis. An example may be a primitive rectangular lattice with lattice parameters a and b where n is the electron density treating the atoms as points. We can define a **reciprocal lattice** made of the points

$$(2\pi \frac{px}{a}, 2\pi \frac{qy}{b}) \quad (4.2.9)$$

for $p, q \in \mathbb{Z}$.

For a crystal that isn't periodic in the same way we cannot use a Fourier series quite so easily. Instead we define a lattice in reciprocal space. We define

$$\mathbf{a}_1^* = 2\pi \frac{\mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)} = \frac{2\pi}{V} \mathbf{a}_2 \times \mathbf{a}_3, \quad (4.2.10)$$

$$\mathbf{a}_2^* = 2\pi \frac{\mathbf{a}_3 \times \mathbf{a}_1}{\mathbf{a}_2 \cdot (\mathbf{a}_3 \times \mathbf{a}_1)} = \frac{2\pi}{V} \mathbf{a}_3 \times \mathbf{a}_2, \quad (4.2.11)$$

$$\mathbf{a}_3^* = 2\pi \frac{\mathbf{a}_1 \times \mathbf{a}_2}{\mathbf{a}_3 \cdot (\mathbf{a}_1 \times \mathbf{a}_2)} = \frac{2\pi}{V} \mathbf{a}_1 \times \mathbf{a}_2. \quad (4.2.12)$$

These are called the **reciprocal lattice vectors**. We then form a lattice from the points

$$\mathbf{G} = v_1 \mathbf{a}_1^* + v_2 \mathbf{a}_2^* + v_3 \mathbf{a}_3^* \quad (4.2.13)$$

with $v_i \in \mathbb{Z}$.

Notice that from this definition

$$\mathbf{a}_i^* \cdot \mathbf{a}_j = 2\pi\delta_{ij}. \quad (4.2.14)$$

We call the space described by \mathbf{a}_i “real space”, “crystal space”, or “direct space”, and the space described by \mathbf{a}_i^* “reciprocal space”.

Reciprocal lattice vectors have units of inverse length.

Notation 4.2.15 — Lattice Vector Conventions There are a few common conventions for what lattice vectors, and reciprocal lattice vectors, are called:

This course

Real space:	$\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$
Reciprocal space:	$\mathbf{a}_1^*, \mathbf{a}_2^*, \mathbf{a}_3^*$

Kittel (textbook)

Real space:	$\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$
Reciprocal space:	$\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$

Crystallographer

Real space:	$\mathbf{a}, \mathbf{b}, \mathbf{c}$
Reciprocal space:	$\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*$

Every crystal therefore has two attached lattices. First, the real-space lattice, in which translation vectors, \mathbf{a}_i , have dimensions of length. This lattice tells us about the periodicity of the atomic arrangement. Second, the reciprocal lattice, in which translation vectors, \mathbf{a}_i^* , have dimensions of inverse length. This lattice is a lattice in Fourier space and is useful for dealing with periodic functions, such as charge density, and scattering processes.

All non-centred lattices have the same form in reciprocal space, for example a cubic lattice will have a cubic reciprocal lattice. Centred lattices aren't so simple.

We can use the reciprocal lattice to define a Fourier series expansion for $n(\mathbf{r})$:

$$n(\mathbf{r}) = \sum_{\mathbf{G}} = n_G e^{i\mathbf{G} \cdot \mathbf{r}} \quad (4.2.16)$$

where $\mathbf{G} = v_1 \mathbf{a}_1^* + v_2 \mathbf{a}_2^* + v_3 \mathbf{a}_3^*$ with $v_i \in \mathbb{Z}$. The Fourier coefficients are given by

$$n_G = \frac{1}{V_{\text{cell}}} \int_{\text{unit cell}} n(\mathbf{r}) e^{-i\mathbf{G} \cdot \mathbf{r}} dV. \quad (4.2.17)$$

Where V_{cell} is the volume of the unit cell.

4.3 The Lattice Vector

The reciprocal lattice vector

$$\mathbf{G}^* = h\mathbf{a}_1^* + k\mathbf{a}_2^* + \ell\mathbf{a}_3^* \quad (4.3.1)$$

is perpendicular to the $(h k \ell)$ plane. To see that this is true we use the fact that this plane intercepts the axes at \mathbf{a}_1/h , \mathbf{a}_2/k , and \mathbf{a}_3/ℓ . This gives us two vectors $\mathbf{a}_1/h - \mathbf{a}_2/k$ and $\mathbf{a}_1/h - \mathbf{a}_3/\ell$ which are in the plane. We can then dot these with \mathbf{G}^* and we see we get zero both times. This result holds for *all* types of crystals. The earlier result that $h\mathbf{a}_1 + k\mathbf{a}_2 + \ell\mathbf{a}_3$ is perpendicular to the plane only holds for orthorhombic crystals where all angles are right angles.

The interplanar spacing, d_{hkl} , is given by

$$d_{hkl} = \frac{2\pi}{|\mathbf{G}^*|}. \quad (4.3.2)$$

To show this is true we notice that consecutive planes cross the \mathbf{a}_1 axis at \mathbf{a}_1/h and $2\mathbf{a}_1/h$ and we use the formula $\hat{\mathbf{n}} \cdot \mathbf{r} = d$, which gives the distance, d , from the origin for a plane with unit normal $\hat{\mathbf{n}}$ containing the point \mathbf{r} . In our case we use \mathbf{r} as the points \mathbf{a}_1/h , and $2\mathbf{a}_1/h$ and $\hat{\mathbf{n}} = \mathbf{G}^*/|\mathbf{G}^*|$. We then work out the two values of d and the difference of these is the interplanar spacing. This works for *all* types of crystals.

For the special case of orthorhombic crystals (all angles right angles) we can simplify this relationship to get

$$\frac{1}{d_{hkl}} = \sqrt{\frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{\ell^2}{c^2}}. \quad (4.3.3)$$

For a cubic crystal (also all sides equal) this simplifies further to

$$\frac{1}{d_{hkl}} = \sqrt{\frac{h^2 + k^2 + \ell^2}{a^2}}. \quad (4.3.4)$$

4.4 Diffraction Intensities

We shall focus on x-ray diffraction as the most common case but similar work should apply to other types of diffraction. In x-ray diffraction electrons absorb x-rays and oscillate. They then emit x-rays in all directions (although not with equal intensity). We assume that the electron density distribution, $n(\mathbf{r})$, is known for a given crystal. We then integrate over all contributions to the emitted radiation, accounting for phase differences. We will find that most cases cancel out leaving us with only a few diffraction angles at which we get bright spots.

We said previously that the scattering amplitude is

$$F(\mathbf{k}, \mathbf{k}') = \int_{\text{crystal}} n(\mathbf{r}) \exp[i(\mathbf{k} - \mathbf{k}') \cdot \mathbf{r}] dV. \quad (4.4.1)$$

Introducing the **scattering vector** $\Delta\mathbf{k}$, sometimes denoted \mathbf{Q} instead, this becomes

$$F_{\Delta\mathbf{k}} = \int_{\text{crystal}} n(\mathbf{r}) e^{-i\Delta\mathbf{k} \cdot \mathbf{r}} dV. \quad (4.4.2)$$

We can recognise this as simply the Fourier transform of the electron density, $n(\mathbf{r})$.

Inserting the Fourier expansion of $n(\mathbf{r})$ we have

$$F_{\Delta k} = \int_{\text{crystal}} \sum_G n_G e^{-(G - \Delta k) \cdot \mathbf{r}} dV \quad (4.4.3)$$

$$\sum_G n_G \int_{\text{crystal}} e^{-(G - \Delta k) \cdot \mathbf{r}} \quad (4.4.4)$$

where we have assumed that the Fourier series converges uniformly and so swapping the sum and integral is valid¹.

Suppose that $G = \Delta k$ for some specific G . Then for this term in the series we have

$$n_G \int_{\text{crystal}} dV = n_G V \quad (4.4.5)$$

where V is the volume of the crystal.

Now suppose $G \neq \Delta k$. We then have that $|G - \Delta k| \approx 2\pi/1 \text{ \AA} \approx 1 \times 10^{11} \text{ m}^{-1}$ and for a typical crystal size of a few millimetres we have $|\mathbf{r}| \approx 1 \text{ mm} \approx 1 \times 10^{-3} \text{ m}$. Hence we find that $|(G - \Delta k) \cdot \mathbf{r}| \approx 1 \times 10^8$. This is huge. The complex exponential factor is then rapidly oscillating. The result is that over the entire integral these oscillations cancel out and we get zero.

Combining these two results we have

$$F_{\Delta k} = \begin{cases} n_G V, & \text{if } \Delta k = G \text{ for some specific } G, \\ 0, & \text{else.} \end{cases} \quad (4.4.6)$$

As a result we only consider the case when $\Delta k = G$ for some $G = h\mathbf{a}_1^* + k\mathbf{a}_2^* + \ell\mathbf{a}_3^*$, which means we only consider F_G , or in alternative notation F_{hkl} . This gives rise to the **Laue diffraction condition** that $F_{\Delta k} \neq 0$ only if $\Delta k = G$.

The diffraction intensity from the plane $(h k \ell)$ is I_{hkl} and this is proportional to $|F_{hkl}|^2 = |F_G|^2 \propto |n_G|$.

There is an equivalent ways to state the Laue diffraction condition. First, rearranging the condition that $G = \Delta k = \mathbf{k}' - \mathbf{k}$ we have $\mathbf{k}' = \mathbf{k} + G$ and so

$$\mathbf{k}' \cdot \mathbf{k}' = (\mathbf{k} + G) \cdot (\mathbf{k} + G) \implies |\mathbf{k}'|^2 = |\mathbf{k}|^2 + 2\mathbf{k} \cdot G + |G|^2. \quad (4.4.7)$$

This gives $-2\mathbf{k} \cdot G = |G|^2$ since $|\mathbf{k}| = |\mathbf{k}'|$ for elastic scattering. We can then consider a reciprocal lattice vector $G' = -G$ in which case we have $2\mathbf{k} \cdot G' = |G'|^2 = G^2$ noticing that G and G' are equivalent when we sum over all lattice vectors we can simply call $G' G$ and we have the condition

$$\mathbf{k} \cdot G = \frac{1}{2}|G|^2. \quad (4.4.8)$$

From this we can define ϑ_{hkl} as the angle between \mathbf{k} and G and we have

$$\mathbf{k} \cdot G = kG \sin \vartheta_{hkl} = \frac{1}{2}|G|^2 = \frac{1}{2} \left(\frac{2\pi}{d_{hkl}} \right)^2. \quad (4.4.9)$$

Noticing that $k = 2\pi/\lambda$ and $G = 2\pi/d_{hkl}$ this becomes

$$2d_{hkl} \sin \vartheta_{hkl} = \lambda \quad (4.4.10)$$

and so we have derived **Bragg's law** without resorting to assumptions of specular reflection.

¹all reasonable physical functions, including the electron density, are smooth enough for uniform convergence of their Fourier series

4.5 Form Factors

The diffraction intensity, I_{hkl} , is proportional to the magnitude squared of the Fourier coefficients of the electron density. Therefore we don't have enough information from the diffraction intensity to recover the Fourier coefficients, and hence to recover the electron density.

In the case when we do have diffraction, i.e. $\Delta\mathbf{k} = \mathbf{G}$, we have the scattering amplitude

$$F_G = \int_{\text{crystal}} n(\mathbf{r}) e^{-i\mathbf{G}\cdot\mathbf{r}} dV. \quad (4.5.1)$$

If our crystal consists of N unit cells then this is

$$F_G = N \int_{\text{unit cell}} n(\mathbf{r}) e^{-i\mathbf{G}\cdot\mathbf{r}} dV. \quad (4.5.2)$$

We can approximate $n(\mathbf{r})$ as a superposition of free-atom electron densities by neglecting the few electrons that take part in the chemical bonding. Doing so and summing over the s atoms in the unit cell we have

$$n(\mathbf{r}) \approx \sum_{j=1}^s n_j(\mathbf{r} - \mathbf{r}_j) \quad (4.5.3)$$

where $n_j(\rho)$ is the free-atom electron density of the j th atom with ρ being the position taking the centre of the atom, \mathbf{r}_j , to be the origin.

We can define the **atomic form factor**, $f_j(G)$, to be

$$f_j(G) := \int n_j(\rho) e^{-i\mathbf{G}\cdot\rho} dV \quad (4.5.4)$$

where we note that the spherical symmetry of atoms means this is a function of G rather than \mathbf{G} . The Fourier coefficients are then given by

$$F_G = N \sum_{j=1}^s f_j(G) e^{-i\mathbf{G}\cdot\mathbf{r}_j}. \quad (4.5.5)$$

In practice it isn't possible to know the exact number of unit cells present in a crystal sample, and we don't need to since we are only interested in proportionality. We then define the **structure factor**

$$S_{hkl} = S_G = \sum_{j=1}^s f_j(G) e^{-i\mathbf{G}\cdot\mathbf{r}_j} = \sum_j f_j \exp[-2\pi i(hx_j + ky_j + \ell z_j)] \quad (4.5.6)$$

where we take $\mathbf{r}_j = x_j \mathbf{a}_1 + y_j \mathbf{a}_2 + z_j \mathbf{a}_3$ and $\mathbf{G} = h\mathbf{a}_1^* + k\mathbf{a}_2^* + \ell\mathbf{a}_3^*$ and used $\mathbf{a}_i^* \cdot \mathbf{a}_j = 2\pi\delta_{ij}$.

It is the structure factor with which we work when considering diffraction. We know that

$$I_{hkl} \propto |S_{hkl}|^2 \quad (4.5.7)$$

and so we lose information about phase. This means that we can't work backwards to find the electron density but this is enough to rule out many crystal structures that are completely wrong.

4.6 Diffraction Techniques

4.6.1 Laue Diffraction

Laue diffraction was the original technique for x-ray crystallography. In this process “white” (meaning a range of wavelengths) broadband spectrum x-rays are used. A single crystal is used as a sample and is placed in the x-ray beam. The diffracted x-rays are collected using some sort of sensor all at once.

This method is best suited to simple crystals, in particular it is often used to ascertain the orientation of a crystal before a more sensitive experiment. It is difficult to model the exact intensities of diffraction due to the range of wavelengths.

4.6.2 Monochromatic Single Crystal Diffraction

This is the modern standard for x-ray crystallography. A single crystal is placed in a beam of monochromatic (single wavelength) x-rays. The diffracted x-rays are collected as the sample is rotated. At any one time x-rays are only diffracted to a few points but by rotating the sample a larger picture can be built up. The result is a picture which is mostly dark but with a few bright spots. From this it is possible to work out which planes caused the diffraction and from the relative intensities we can measure the magnitude of the structure factor.

For less sensitive work x-ray tubes can produce the required x-ray beams. More sensitive work requiring higher energy, more focused beams often uses synchrotron radiation.

4.6.3 Monochromatic Powdered Crystal Diffraction

Often it is not possible to come up with a single crystal large enough for the procedure in the previous section. Instead a polycrystalline powder can be used. The setup is otherwise identical. Since the crystals in the powder are arranged randomly it is not necessary to rotate the sample. A single image can be captured. The result will be a series of rings as the crystals diffract in random directions but always at the same angle. Integrating around the rings gives the relative intensity, it is not enough to consider the intensity at a point on the ring due to the varying size of the rings. It should be possible to assign to each ring the plane from which the diffraction occurred.

4.6.4 Atomic Form Factor

Diffraction experiments, particularly powder diffraction, will show an overall decrease in intensity with increasing diffraction angle. Partly this is geometric, at larger angles the rings are larger and so the same intensity would be spread over a larger area. However, the effect remains when taking this into account.

A second factor is destructive interference of x-rays scattered from different parts of the same atom. The atomic form factor is the Fourier transform of the charge distribution. It describes the interference of waves scattered from different parts of the same atom. When we consider x-rays the atomic form factor $f(0)$ gives the number of electrons in the atom, which can be seen from

$$f(0) = \int n(\rho) e^0 dV. \quad (4.6.1)$$

Here ρ is the distance from the centre of the atom.

If we take the atoms to be point like scatterers, which is almost the case for neutron diffraction, then the atomic form factors reduce to a constant.

The atomic form factor is a function of G , which, combining Bragg's law and $d = 2\pi/G$ is given by $G = 4\pi \sin(\theta)/\lambda$. The values $f(\sin(\theta)/\lambda)$ are tabulated for all elements.

4.6.5 Neutron Diffraction

The data produced by neutron diffraction is very similar to that produced by x-ray diffraction. The main difference being that x-rays are scattered by the extended electron cloud whereas neutrons are scattered by the nucleus, which is essentially a point particle. This results in atomic form factors which are independent of the scattering angle. We call these form factors the neutron scattering length. The result is that neutrons scattering doesn't have the same drop off with diffraction angle.

Neutrons are generally produced in one of two ways. The more old fashioned way is from nuclear reactors. More modern setups us spallation sources, which is a fancy way of saying we smash protons into targets at high energy to produce neutrons. One issue with neutron diffraction is that neutrons are neutral and therefore we can't use magnetic fields to accelerate them and focus beams in the same way we do with other particles.

4.7 Systematic Absences

After a diffraction experiment if we have successfully assigned planes to all data points we may see that some planes don't appear. These are called **systematic absences**. These are a result of terms cancelling each other out due to symmetry. This is best demonstrated with an example.

4.7.1 BCC Systematic Absences

Consider a BCC crystal. The structure factor is

$$S_{hkl} = S_G = \sum_j f_j e^{-2\pi i (hx_j + ky_j + \ell z_j)}. \quad (4.7.1)$$

We consider the conventional cubic lattice, not the primitive lattice. The lattice vectors are $\mathbf{a}_i = a\mathbf{e}_i$. The basis has atoms positioned at $\mathbf{r}_1 = (0, 0, 0)$, and $\mathbf{r}_2 = (1/2, 1/2, 1/2)$. There is only one type of atom in this simple example and so $f_1 = f_2 = f$. Putting this into the formula for the structure factor we have

$$S_{hkl} = f [e^{-2\pi i (0)} + e^{-2\pi i (h+k+\ell)/2}] = f [1 + e^{i\pi(h+k+\ell)}]. \quad (4.7.2)$$

We see that when $h + k + \ell$ is odd the second exponential gives -1 and therefore the structure factor is zero:

$$S_{hkl} = \begin{cases} 2f, & h + k + \ell \text{ even}, \\ 0, & h + k + \ell \text{ odd}. \end{cases} \quad (4.7.3)$$

So there is a systematic absence when $h + k + \ell$ is odd.

We can see this as a result of the symmetry of the crystal, hence why we cannot predict this effect from the lower symmetry primitive cell. We find a similar result for an FCC lattice:

$$S_{hkl} = \begin{cases} 2f, & h, k, \ell \text{ all odd, or all even}, \\ 0, & \text{else}. \end{cases} \quad (4.7.4)$$

As another example in a C-centre lattice² we have a systematic absence when $h + k$ is even.

²recall that this is a lattice with a lattice point at the centre of the C face, that is the face spanned by \mathbf{a} and \mathbf{b} .

4.8 FCC and BCC Reciprocal Lattices

Now that we are discussing diffraction from BCC and FCC crystals while we need to consider the reciprocal lattices for BCC and FCC. Recall that non-centred lattices are their own reciprocals, albeit with different lattice parameters.

Consider the primitive cell for an FCC crystal. The primitive translation vectors are

$$\mathbf{a}_1 = \frac{a}{2}(\hat{\mathbf{y}} + \hat{\mathbf{z}}), \quad \mathbf{a}_2 = \frac{a}{2}(\hat{\mathbf{z}} + \hat{\mathbf{x}}), \quad \text{and} \quad \mathbf{a}_3 = \frac{a}{2}(\hat{\mathbf{x}} + \hat{\mathbf{y}}). \quad (4.8.1)$$

The volume of the primitive cell is

$$V = |\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)| = \frac{1}{4}a^3. \quad (4.8.2)$$

The reciprocal translation vectors are then given by

$$\mathbf{a}_1^* = \frac{2\pi}{V}\mathbf{a}_2 \times \mathbf{a}_3 = \frac{2\pi a^2}{4V}(\hat{\mathbf{x}} + \hat{\mathbf{z}}) \times (\hat{\mathbf{x}} + \hat{\mathbf{y}}) = \frac{2\pi}{a}(\hat{\mathbf{z}} + \hat{\mathbf{y}} - \hat{\mathbf{x}}) \quad (4.8.3)$$

and similarly

$$\mathbf{a}_2^* = \frac{2\pi}{a}(\hat{\mathbf{x}} + \hat{\mathbf{z}} - \hat{\mathbf{y}}), \quad \text{and} \quad \mathbf{a}_3^* = \frac{2\pi}{a}(\hat{\mathbf{y}} + \hat{\mathbf{x}} - \hat{\mathbf{z}}). \quad (4.8.4)$$

These are simply the translation vectors for some BCC lattice in reciprocal space. We see that the reciprocal lattice for a FCC lattice is the BCC lattice, and since the reciprocal of the reciprocal is (up to a constant) the original lattice we know also that the reciprocal of a BCC lattice is the FCC lattice.

4.9 Diffraction Beyond Simple Systems

4.9.1 DNA

Perhaps the most important example of x-ray diffraction on a complex system was the determination of the double-helix structure of DNA. This was discovered in experiments conducted by Rosalind Franklin. However, she was not fully recognised for this work at the time since she was a woman. She died before the Nobel prize was awarded to Francis Crick, James Watson, and Maurice Wilkins for their contributions to the discovery of the double helix and it is widely believed that had she been alive she would have been included in this prize.

The key use of x-ray crystallography was to rule out a triple-helix structure that had been suggested. Like most x-ray crystallography it wasn't possible to confirm the double-helix structure through x-ray crystallography alone, but it could rule out competing hypotheses.

4.9.2 Small Angle Scattering

Small angle scattering is a technique used to obtain information about structures with characteristic length in the range 1 nm to 100 nm. This covers structures from small viruses down to individual molecules. As the name suggests diffraction patterns at low angles, in the range 0.1° – 10° . Both x-rays and neutrons can be used, in which case the process is called small angle x-ray scattering (SAXS) or small angle neutron scattering (SANS), respectively. One advantage of SANS is that neutrons interact more with hydrogen than x-rays do and for biological materials we are often interested in the location of hydrogen atoms.

We saw with crystals that the atomic form factor plays an important role. This idea generalises to the notion of a form factor which encodes information about the size and shape of the object being studied. In small angle scattering it is conventional to work with the quantity $Q = 4\pi \sin(\vartheta)/\lambda$, instead of the diffraction angle, 2ϑ .

4.9.3 Aperiodic Crystals

So far we have only considered periodic crystals, and we will consider to do so for most of the course. We would be remiss, however, not to at least mention aperiodic crystals.

It was initially believed that there are no crystals with five-fold symmetry. This is related to the idea that regular pentagons cannot tile the plane. However, (with the modern definition of crystals based on an essentially sharp diffraction pattern) there have been crystals discovered with 10-fold rotationally symmetric diffraction patterns, and therefore they also necessarily have five-fold rotational symmetry. The first example being a crystal made of zinc, magnesium and holmium. The existence of such a crystal is related to the ability to tile the plane aperiodically with five-fold symmetry, the most famous example using Penrose tiles.

Aperiodic crystals are often called **quasicrystals**, since they don't fit earlier definitions of crystals. Since their first discovery many other examples have been found.

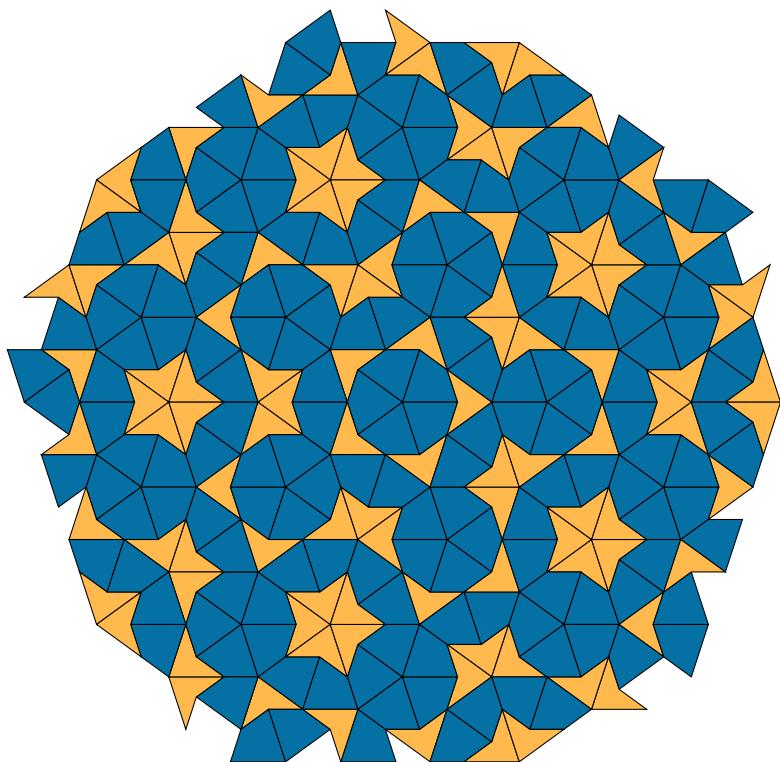


Figure 4.2: An aperiodic tiling of the plane with five-fold symmetry, known as a Penrose tiling.

Five

Crystal Binding

So far we've considered the shape of various crystals, but not why they form these shapes. To answer this we need to understand what holds crystals together. It's then a case of finding a minimum energy structure and over time a crystal will tend towards this form. There are four types of bonding that we need to consider. These are

1. Van der Waals' forces
2. Ionic bonding
3. Covalent bonding
4. Metallic bonding

Real crystals usually have a mixture of these. Van der Waals forces are the weakest and appear predominantly in noble gas crystals, since they are vastly overshadowed by any other bonding. Ionic bonds form between metals and non-metals when they exchange electrons to form ions which then attract each other. Covalent bonds form between non-metals when they share electrons, which they are both attracted to. Metallic bonds form, obviously, between metals where they lose electrons to form a sea of electrons and the resulting positive ions are then bound to this sea.

When measuring bond strength force is not a very good measure since the force depends on the distance. Instead a better measure is the change in energy between a free system, with no interactions, and the system with interactions. A negative change in energy means that the interaction is attractive as the bound state is energetically favourable. On the other hand, a positive energy change means that the interaction is repulsive.

Typically binding energies are

1. ~ 0.1 eV per atom for rare gas crystals,
2. ~ 1 eV per atom for alkali metals,
3. ~ 1 eV/atom to 10 eV/atom for other elemental crystals, and
4. ~ 2 eV/atom to 6 eV/atom for ionic crystals.

These are the **cohesive energies** of the crystals. They are defined as the energy required to form separated neutral atoms from the crystal. Precisely this requires moving the atoms infinitely far apart but in practice almost all of the energy is needed for just a small separation.

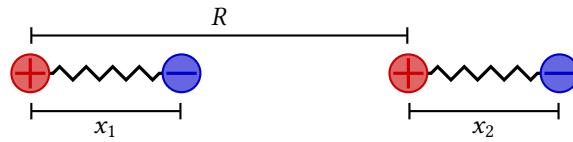


Figure 5.1: The two dipole oscillators used to model van der Waals interactions.

5.1 Van der Waals

5.1.1 Attractive Component

The **van der Waals** interaction is important in weakly bound crystals, such as noble-gas crystals, and also in soft matter, for example, when considering colloidal suspensions. They can even have microscopic effects, such as allowing Gecko's feet to stick to walls. Van der Waals interactions are present in all crystals but other interactions are much stronger if they are present.

The mechanism behind the attractive component of the van der Waals interaction is instantaneous dipoles. In general we can model atoms as two charges, a positive charge fixed at a nucleus and a negative charge at the centre of the electron cloud. This negative charge is not stationary but moves around slightly due to random fluctuations in charge density. The charge separation results in an instantaneous dipole. If there is a nearby atom then this will induce in the atom a dipole aligned such that the two atoms attract each other.

Of course this atom also has random fluctuations in charge density and so is forming its own instantaneous dipoles and interacting with its neighbours and so on.

5.1.1.1 Oscillator Model

We are interested in the attractive interaction between two dipoles. We can model these as quantum harmonic oscillators, see Figure 5.1. Each formed from two charges, $\pm e$, which represent the charge of the nucleus and the valence electron. These are connected by a spring, which models the attraction between the charges. We fix the positive charges at some length R , which is the nuclear separation. Each pair of charges is separated by x_i . We want to know how the energy of the system changes as the dipoles are brought closer together. We will see that the solution is $E \propto R^{-6}$.

We start by considering the unperturbed system. This is equivalent to considering the case of $R = \infty$. The Hamiltonian for the unperturbed system is

$$\mathcal{H}_0 = \frac{p_1^2}{2m} + \frac{1}{2}Cx_1^2 + \frac{p_2^2}{2m} + \frac{1}{2}Cx_2^2. \quad (5.1.1)$$

This is simply the sum of the two independent oscillator's Hamiltonians. The natural frequency of these oscillators, ω_0 , relates to the constant C by $C = m\omega_0^2$. The energy of a single oscillator in the n th excited state is

$$E_n = \left(n + \frac{1}{2} \right) \hbar\omega_0 \quad (5.1.2)$$

with¹ $n \in \mathbb{N}$. It follows that the zero-point energy (ZPE) of the two oscillator system is

$$E_0 = 2 \cdot \frac{1}{2} \hbar\omega_0. \quad (5.1.3)$$

¹ $\mathbb{N} := \{0, 1, 2, \dots\}$

We can now add in a perturbation to the Hamiltonian corresponding to the interactions between the charges. The perturbation Hamiltonian is simply the sum of the Coulomb interactions between the pairs of charged particles:

$$\mathcal{H}_1 = \frac{e^2}{R} + \frac{e^2}{R+x_1-x_2} - \frac{e^2}{R+x_1} - \frac{e^2}{R-x_2}. \quad \text{CGS} \quad (5.1.4)$$

R

This equation is in CGS units, we will flag equations in CGS units with CGS throughout. The CGS system of units uses centimetres, grams and seconds as base units, as opposed to the SI, which uses metres, kilograms, and seconds. To convert this equation from CGS to SI replace e^2 with $e^2/(4\pi\epsilon_0)$.

Returning to the problem at hand, this Hamiltonian is not easy to deal with. Assuming that $|x_1|, |x_2| \ll R$ we can combine the terms over a common denominator and drop all but the highest order term. This leaves us with

$$\mathcal{H}_1 \approx -\frac{2e^2x_1x_2}{R^3}. \quad (5.1.5)$$

We can now consider the Hamiltonian for the full system:

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1 = \frac{p_1^2}{2m} + \frac{1}{2}Cx_1^2 + \frac{p_2^2}{2m} + \frac{1}{2}Cx_2^2 - \frac{2e^2x_1x_2}{R^3}. \quad \text{CGS} \quad (5.1.6)$$

This is easiest to work with if we diagonalise. To do this we introduce a normal mode transformation defining

$$x_s := \frac{1}{\sqrt{2}}(x_1 + x_2), \quad \text{and} \quad x_a := \frac{1}{\sqrt{2}}(x_1 - x_2). \quad (5.1.7)$$

Similarly we define the momenta

$$p_s := \frac{1}{\sqrt{2}}(p_1 + p_2), \quad \text{and} \quad p_a := \frac{1}{\sqrt{2}}(p_1 - p_2). \quad (5.1.8)$$

We can rearrange these to find x_i and p_i :

$$x_1 = \frac{1}{\sqrt{2}}(x_s + x_a), \quad x_2 = \frac{1}{\sqrt{2}}(x_s - x_a), \quad (5.1.9)$$

$$p_1 = \frac{1}{\sqrt{2}}(p_s + p_a), \quad p_2 = \frac{1}{\sqrt{2}}(p_s - p_a). \quad (5.1.10)$$

Substituting these into the Hamiltonian and simplifying gives

$$\mathcal{H} = \underbrace{\frac{p_s^2}{2m} + \frac{1}{2} \left(C - \frac{2e^2}{R^3} \right) x_s^2}_{=: m\omega_s^2} + \underbrace{\frac{p_a^2}{2m} + \frac{1}{2} \left(C + \frac{2e^2}{R^3} \right) x_a^2}_{=: m\omega_a^2}. \quad \text{CGS} \quad (5.1.11)$$

We see that this is the sum of two independent oscillators with frequencies

$$\omega_{a,s} = \sqrt{\frac{1}{m} \left(C \pm \frac{2e^2}{R^3} \right)} = \omega_0 \sqrt{1 \pm \frac{2e^2}{CR^3}}. \quad \text{CGS} \quad (5.1.12)$$

Using the Taylor expansion, $\sqrt{1+\varepsilon} \approx 1 + x/2 - x^2/8$, this becomes

$$\omega_{a,s} \approx \omega_0 \left[1 \pm \frac{1}{2} \frac{2e^2}{CR^3} - \frac{1}{8} \left(\frac{2e^2}{CR^3} \right)^2 \right]. \quad \text{CGS} \quad (5.1.13)$$

The total zero-point energy of the coupled system is then

$$E'_0 = \frac{1}{2} \hbar \omega_s + \frac{1}{2} \hbar \omega_a = \hbar \omega_0 \left[1 - \frac{1}{8} \left(\frac{2e^2}{CR^3} \right)^2 \right] = \hbar \omega_0 \left(1 - \frac{A}{R^6} \right) \quad \text{CGS} \quad (5.1.14)$$

where

$$A = \frac{1}{8} \left(\frac{2e^2}{C} \right)^2. \quad \text{CGS} \quad (5.1.15)$$

Recall that the zero point energy of the uncoupled system is $E_0 = \hbar \omega_0$ and we see that the reduction in energy due to the coupling is

$$\Delta U = E'_0 - E_0 = -\frac{A}{R^6}. \quad (5.1.16)$$

This is negative and hence corresponds to an attractive interaction. The binding energy varies as R^{-6} and the attractive force,

$$F = -\frac{d\Delta U}{dR}, \quad (5.1.17)$$

varies as R^{-7} .

This attractive interaction depends on the zero-point energy of the system and hence is an inherently quantum effect.

This interaction is known under a few names. We have been calling it the van der Waals interaction, after Johannes van der Waals, the experimentalist who discovered the effect. It is also sometimes called the **London interaction**, after Fritz London, the theorist who first explained it. At the time of solving this problem London was working on the dispersion of light in glass and the relation of wave length and index of refraction, this can also be explained by the oscillator model of an atom. See the notes from the optics part of the Electromagnetism course. For this reason these are also sometimes called the **dispersion interaction**. The mechanism of inducing dipoles means that these are also called **induced dipole-dipole interactions**.

5.1.2 Repulsive Term

The van der Waals interaction is entirely attractive. Yet, atoms don't all collapse to a single point. There must be some repulsive force we have not accounted for. It would not be unreasonable to assume that this comes from the Coulomb interaction between the positively charged nuclei. However, there is a far greater source of repulsion, from the Pauli exclusion principle.

As the two nuclei are brought together their wave functions start to overlap. As a result the Pauli exclusion principle applies to *all* electrons in the combined two-atom system. Assuming that both atoms start in their ground states the only way to combine the wave functions is to have some electrons become excited so that they aren't in the same state. This requires energy, meaning the potential energy increases, and hence the result is a repulsive force.

To see an example of this consider combining two hydrogen atoms with their electrons both in the ground state, 1s, but with opposite spins. When combined we get a ${}^2\text{He}$ atom with two electrons, both in the ground state 1s and with opposite spins. The total electron energy is -78.98 eV . A similar situation with two electrons with aligned spins can only occur if upon merging one of the electrons is excited to 2s (note that the spin can't flip as we must conserve angular momentum). The total electron energy is then -59.38 eV . This is an increase in energy of -19.6 eV . For comparison in hydrogen the electron is only bound with -13.6 eV in its ground state.

We can capture the essential features of this repulsive effect with an empirical potential. These are potentials which we use through no justification other than they work. One common potential is

$$U(R) = \lambda e^{-R/\rho} \quad (5.1.18)$$

where R is the internuclear separation, and λ and ρ are constants of some characteristic energy and length scale. A simpler potential which we will make use of is

$$U(r) = \frac{B}{R^{12}} \quad (5.1.19)$$

where B is some constant. There is nothing particularly important about the exponent being 12. Models with R^{-10} , for example, would work just as well with a different constant. The choice of 12 is mostly for computational efficiency, especially in the pre-computer days. Once we have calculated R^{-6} for the attractive component computing R^{-12} is as simple as squaring this result.

5.1.3 Lennard-Jones Potential

The **Lennard-Jones potential** is the result of combining both the attractive and repulsive interactions. It is

$$U_{\text{LJ}}(R) = 4\epsilon \left[\left(\frac{\sigma}{R}\right)^{12} - \left(\frac{\sigma}{R}\right)^6 \right] \quad (5.1.20)$$

where ϵ and σ are parameters related to the energy and length scale respectively. This describes the interaction between a *single pair* of atoms.

The minimum of the potential, that is the equilibrium interatomic distance for a diatomic Lennard-Jones molecule, is found by differentiation:

$$0 = \frac{dU_{\text{LJ}}}{dR} \Big|_{R=r_0} \quad (5.1.21)$$

$$= 4\epsilon \left[-12 \frac{\sigma^{12}}{r_0^{13}} + 6 \frac{\sigma^6}{r_0^7} \right] \quad (5.1.22)$$

$$\implies 12 \frac{\sigma^{12}}{r_0^{13}} = 6 \frac{\sigma^6}{r_0^7} \quad (5.1.23)$$

$$\implies \frac{\sigma^6}{r_0^6} = \frac{1}{2} \quad (5.1.24)$$

$$\implies \frac{r_0}{\sigma} = 2^{1/6} \approx 1.12246. \quad (5.1.25)$$

5.2 Van der Waals Crystal

We now have the Lennard-Jones potential, describing the interaction between *pairs* of atoms due to van der Waals interactions and the repulsive effect of Pauli's exclusion principle. To consider an entire crystal we need to consider all pairwise interactions between atoms. Start by numbering all atoms to give them names. We take some reference atom, i , and we sum over all atoms, j , apart from i . For each term of the sum we calculate the potential, $U_{\text{LJ}}(r_{ij})$, where r_{ij} is the distance between atom i and atom j . This calculates all pairwise interactions with atom i . If we do this for every atom one at a time taking the role of the reference atom then the result will be *twice* the total potential of the entire crystal. The result is two times too large since we have double counted the interaction between i and j and the interaction between j and i . For this reason we add in a factor of $1/2$. In maths the process we've just described is

$$U_{\text{tot}} = \frac{1}{2} \sum_{i=1}^N \sum_{j \neq i} U_{\text{LJ}}(r_{ij}). \quad (5.2.1)$$

Assuming that N is large and the potential drops away with distance only the nearby atoms are important. The ordered nature of the crystal means that most atoms are in a similar configuration of nearby atoms. The exception is for atoms on or near the surface of the crystal. However, the number of surface atoms goes with the square of the crystal size and the number of interior atoms goes with the cube. Therefore for all but the smallest crystals we can treat the surface atoms as interior atoms and the error will be relatively small. The result is that each term of the sum over i is approximately the same and so we make the approximation

$$U_{\text{tot}} = \frac{N}{2} \sum_{j \neq i} U_{\text{LJ}}(r_{ij}) \quad (5.2.2)$$

where we now choose some random interior atom to be i .

Substituting in the full potential we have

$$U_{\text{tot}} = \frac{N}{2} 4\epsilon \left[\sum_{j \neq i} \left(\frac{\sigma}{r_{ij}} \right)^{12} - \sum_{j \neq i} \left(\frac{\sigma}{r_{ij}} \right)^6 \right]. \quad (5.2.3)$$

We want to come up with a formula that we can apply to any lattice. For this reason we define $r_{ij} = Rp_{ij}$ where R is the minimum interatomic spacing in the crystal, that is the nearest neighbour distance. This allows us to factor out the crystal-specific information from the sums leaving us with

$$U_{\text{tot}}(R) = \frac{N}{2} 4\epsilon \left[\underbrace{\sum_{j \neq i} \frac{1}{p_{ij}^{12}}}_{=:a_{12}} - \underbrace{\sum_{j \neq i} \frac{1}{p_{ij}^6}}_{=:a_6} \right]. \quad (5.2.4)$$

Here we have defined the **lattice sums**, a_6 and a_{12} :

$$a_n := \sum_{j \neq i} \frac{1}{p_{ij}^n}. \quad (5.2.5)$$

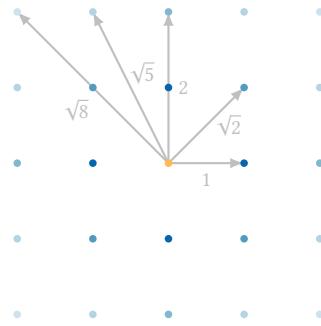


Figure 5.2: A square crystal with distances from some reference atom.

5.2.1 Lattice Sums

5.2.1.1 Two Dimensions

We will start with an example of a lattice sum in two dimensions, as these are easier to visualise. For geometrical simplicity we consider a square crystal. Figure 5.2 shows a reference atom and the distances to its five nearest neighbours. We can also count the number of nearest neighbours from this diagram. Since each nearest neighbour of a given order is the same distance from the reference atom we can combine the terms of the sum corresponding to the nearest neighbours. Going up to the fifth nearest neighbours we get have

$$a_{12} = \frac{4}{1^{12}} + \frac{4}{\sqrt{2}^{12}} + \frac{4}{2^{12}} + \frac{8}{\sqrt{5}^{12}} + \frac{4}{\sqrt{8}^{12}} + \dots \quad (5.2.6)$$

$$\approx 4 + 0.0625 + 0.000977 + 0.000512 + \dots \quad (5.2.7)$$

$$\approx 4.064. \quad (5.2.8)$$

From this we can see that the lattice sum is approximately the number of nearest neighbours with a small correction for the second nearest neighbours and an even smaller correction for the third nearest neighbours and so on.

5.2.1.2 Three Dimensions

A similar process can be applied to three-dimensional crystal structures, although the geometry is slightly more complicated. Again, we would find that each lattice sum is essentially the number of nearest neighbours with small corrections, although the corrections are slightly large than the two-dimensional case since the number of nearest neighbours goes with distance squared in the three-dimensional case but only with distance in the two-dimensional case.

Fortunately people have already calculated the lattice sums for common lattices to far more precision than we could reasonably do. They found

$$\text{FCC: } a_{12} = 12.13188, \quad a_6 = 14.45392, \quad (5.2.9)$$

$$\text{HCP: } a_{12} = 12.13229, \quad a_6 = 14.45489, \quad (5.2.10)$$

$$\text{BCC: } a_{12} = 9.11418, \quad a_6 = 12.2533. \quad (5.2.11)$$

Notice that FCC and HCP have almost identical lattice sums. This is because the dominant effect in the lattice terms is the distance to nearest neighbours, which in turn is linked to the packing density, and the packing density of FCC and HCP is

the same. It should also be noted that a_{12} is approximately the number of nearest neighbours, also known as the **coordination number**. Both FCC and HCP have a coordination number of 12. BCC has a coordination number of 8. The discrepancy between $a_{12} \approx 9.1$ and the coordination number 8 is due to the fact that the second nearest neighbours in a BCC crystal aren't that much further away than the nearest neighbours and therefore have a non-negligible contribution to the lattice sum.

5.2.2 Van der Waals Crystal Properties

Returning now to the potential we can find the equilibrium interatomic spacing in a crystal by finding the minimum of the potential:

$$0 = \frac{dU_{\text{tot}}}{dR} \Big|_{R=R_0} \quad (5.2.12)$$

$$= 2N\varepsilon \left[-12a_{12} \frac{\sigma^{12}}{R_0^{13}} + 6a_6 \frac{\sigma^6}{R_0^7} \right] \quad (5.2.13)$$

$$\implies 12a_{12} \frac{\sigma^{12}}{R_0^{13}} = 6a_6 \frac{\sigma^6}{R_0^7} \quad (5.2.14)$$

$$\implies \frac{\sigma^6}{R_0^6} = \frac{a_6}{2a_{12}} \quad (5.2.15)$$

$$\implies \frac{\sigma}{R_0} = \left(\frac{a_6}{2a_{12}} \right)^{1/6}. \quad (5.2.16)$$

For comparison the interatomic spacing between two otherwise free atoms gave us $r_0/\sigma = 2^{1/6} \approx 1.12$. For FCC and HCP we find that $R_0/\sigma \approx 1.09$ and for BCC $R_0/\sigma \approx 1.07$. We see that the interatomic spacing decreases slightly from the free diatomic molecule.

In general the potential energy at equilibrium is

$$U_{\text{tot}}(R_0) = 2N\varepsilon \left[\left(\frac{a_6}{2a_{12}} \right)^{12} \right] a_{12} - \left(\frac{a_6}{2a_{12}} \right) = -4\varepsilon N \frac{a_6^2}{4a_{12}}. \quad (5.2.17)$$

A more useful quantity is the binding energy per atom

$$U_0 = \frac{1}{N} U_{\text{tot}}(R_0) = -4\varepsilon N \frac{a_6^2}{8a_{12}}. \quad (5.2.18)$$

Evaluating this for FCC, HCP, and BCC crystals we find

$$\text{FCC: } -2.1525 \cdot 4\varepsilon, \quad \text{HCP: } -2.1582 \cdot 4\varepsilon, \quad \text{and} \quad \text{BCC: } -2.0592 \cdot 4\varepsilon. \quad (5.2.19)$$

From this we see that FCC and HCP are the most stable van der Waals crystals. The small difference between their binding energies per atom is negligible compared to approximations made in the derivation.

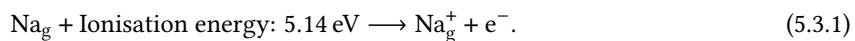
This observation is born out in reality. When we cool the noble gases, Ne, Ar, Kr, and Xe down to a few kelvin they do indeed form FCC crystals. The lightest noble gas, He, doesn't form a solid, even at absolute zero. This is because its zero-point energy is higher than the very small binding energy provided by van der Waals interactions. If we increase the pressure, essentially artificially decreasing interatomic spacing and hence increasing the strength of van der Waals interactions, then He does form a crystal structure and this too agrees with our prediction that FCC and HCP are preferentially selected for by energy minimisation.

5.3 Ionic Binding

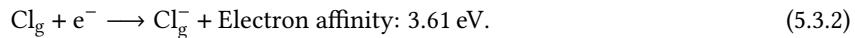
5.3.1 Cohesive Energy

Recall that when we defined cohesive energy at the start of the chapter (Chapter 5) we defined it as the energy required to separate a crystal into its *neutral* constituent atoms. For an ionic crystal this means we need to consider not only the binding energy but also the energy required to create the ions from the neutral atoms. That is the ionisation energy for the positive ions and electron affinity for the negative ions.

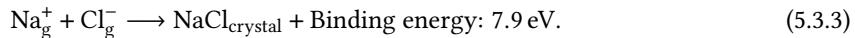
Consider, for example, NaCl. This is formed from Na^+ ions and Cl^- ions. We make the Na^+ ions in the reaction



Here the subscript g denotes that we consider a gas of sodium atoms and ions, this is simply because this allows us to treat the interactions between atoms/ions as negligible. We make the Cl^- ions in the reaction



Note that both energies above are per atom/ion. We then form a single NaCl unit in the reaction



Combining these we see that when creating NaCl from gaseous sodium and chlorine the energy change is

$$5.1 \text{ eV} - 3.6 \text{ eV} - 7.9 \text{ eV} = -6.4 \text{ eV} \quad (5.3.4)$$

per NaCl unit. Hence the cohesive energy of NaCl is 6.4 eV per NaCl unit. We call this the energy per formula unit, sometimes abbreviated as eV/f.u..

In reality in standard conditions chlorine forms a gas but sodium is in a metallic state. However, the cohesive energy of sodium metal is only 1.1 eV/atom, and so creating NaCl from the elements in their normal forms is still energetically favourable.

One of the reasons that the ionic NaCl is more stable than the elemental forms is that in the process of forming the ions both Na and Cl achieve full outer shells, Na goes from $[\text{Ne}]3s$ to $[\text{Ne}]$, and Cl goes from $[\text{Ne}]3s^23p^5$ to $[\text{Ne}]3s^23p^6 = [\text{Ar}]$. These closed-shell configurations are very stable. It would require a large amount of energy to transfer more electrons, for example to create Na^{2+} or Cl^{2-} .

5.3.2 Ionic Crystals

We can now calculate the binding energy of an ionic crystal. The approach is similar to the van der Waals crystal but slightly complicated by having to consider two different ions with different charges. For simplicity we consider crystals formed of two ions of equal and opposite charge, this covers many, but not all, ionic crystals, for example this works for NaCl, but not CaCl₂.

While all crystals, including ionic crystals, have van der Waals interactions they are very weak and so we ignore them in this case.

Recall that the Pauli exclusion principle leads to a short-range repulsive force. We model this with an empirical potential, V_r , which we take to be $V_r(r) = \lambda e^{-r/\rho}$,

another common form is $V_r(r) = A/r^n$. We will leave the result in such a way that we can easily swap out different potentials. Further, we assume that this potential is short-range and we can ignore it between anything but nearest neighbours.

The other interaction we have to consider is the long-range Coulomb interaction:

$$V_C(r) = \pm \frac{q^2}{4\pi\epsilon_0 r}. \quad (5.3.5)$$

Here the ions have charge $\pm q$, and the sign of $\pm q^2$ depends on whether we consider a pair of atoms of equal or opposite charge. This potential decreases slowly with r (at least when compared to V_r). In fact, $V_C \propto 1/r$. The number of k -nearest neighbours increases with r , in two dimensions it increases as r , in three dimensions it increases as r^2 . This causes problems since we can't be sure that a sum over all atoms of the Coulomb potential will converge.

The total energy of the crystal is then

$$U_{\text{tot}} = \frac{1}{2} \sum_i \sum_{j \neq i} U_{ij} = \frac{1}{2} \sum_i \sum_{j \neq i} [V_r(r_{ij}) + V_C(r_{ij})]. \quad (5.3.6)$$

Let $p_{ij} = r_{ij}/R$, where R is the smallest interionic distance in the crystal. Then the interaction potential between a pair of atoms is

$$U_{ij} = \begin{cases} -\frac{1}{4\pi\epsilon_0 R} \frac{q^2}{r_{ij}} + V_r(R), & \text{for nearest neighbours,} \\ \pm \frac{1}{p_{ij}} \frac{q^2}{4\pi\epsilon_0 R}, & \text{otherwise.} \end{cases} \quad (5.3.7)$$

The interaction potential for a single ion is then

$$U_i = zV_r(R) - \frac{q^2}{4\pi\epsilon_0 R} \sum_{j \neq i} \frac{(\pm)}{p_{ij}}. \quad (5.3.8)$$

Here z is the number of nearest neighbours and

$$(\pm) := \begin{cases} +1, & \text{if atoms } i \text{ and } j \text{ have opposite charge,} \\ -1, & \text{if atoms } i \text{ and } j \text{ have the same charge.} \end{cases} \quad (5.3.9)$$

The total energy is then given by

$$U_{\text{tot}} = \frac{1}{2} \sum_i \sum_{j \neq i} U_{ij} = \frac{1}{2} 2N \sum_{j \neq i} = NU_i \quad (5.3.10)$$

where, as with the van der Waals case, we have made the approximation that all atoms are equivalent and hence a sum over all atoms is equivalent to a factor of N .

As with the van der Waals case we have written U_i in such a form that we separate the geometry of the crystal from the specific details of the constituents. The sum in U_i depends only on the geometry of the potential. We call its value the **Madelung constant**, α :

$$\alpha := \sum_{j \neq i} \frac{(\pm)}{p_{ij}}. \quad (5.3.11)$$

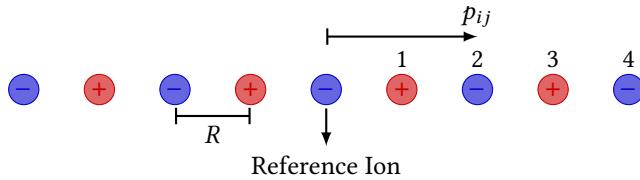


Figure 5.3: One-dimensional crystal of alternating positive and negative ions.

This is equivalent to the lattice sums a_{12} and a_6 of the Lennard-Jones crystal. The values of α have been calculated for many different crystal structures so we can usually just look up the value. Thus the total energy is

$$U_{\text{tot}}(R) = N \left[zV_r(R) - \alpha \frac{q^2}{4\pi\epsilon_0 R} \right] \quad (5.3.12)$$

for N pairs of ions.

We define the **Madelung energy** as

$$-N\alpha \frac{q^2}{4\pi\epsilon_0 R}. \quad (5.3.13)$$

This is the total energy from all ionic Coulomb interactions for N pairs of ions. The Madelung energy must be negative for the crystal to be stable, and hence α must be positive. This is needed so that $U_{\text{tot}}(R)$ has a minimum, otherwise the minimum total energy would be achieved by separating the ions to $R = \infty$.

As an example we will compute the Madelung constant for a one dimensional crystal. We consider an ideal crystal which is an infinite row of ions, alternating positive and negative. See Figure 5.3. By the symmetry of the crystal we only need to consider ions on one side of the reference ion and simply double the result for the whole crystal. For the first ion the distance is 1 and the charge is opposite so the contribution to the Madelung constant is 1. For the second ion the distance is 2 and the charge is the same so the contribution to the Madelung constant is $-1/2$. Continuing on we see that the Madelung constant is

$$\alpha = 2 \left[1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots \right] = 2 \ln 2 \approx 1.386 \quad (5.3.14)$$

where we have identified the Taylor expansion

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots \quad (5.3.15)$$

Computing the Madelung constant in higher dimensions is more tricky. As a start point summing in order of nearest neighbours doesn't work very well since the terms will oscillate from positive to negative. A better method is to sum over spherical shells. The resulting sum is conditionally convergent, and when it does converge it doesn't do so particularly quickly since the Coulomb potential goes as $1/r$ and the number of ions in a spherical shell of radius r goes as r^2 . Fortunately we know that NaCl, and other structures, exist and therefore we can assume that the sum converges.

A better method for calculating the Madelung constant is the Ewald method. It involves transforming the sum to Fourier space for the calculation, and it turns out the convergence is much quicker in Fourier space. This method is used by most modern programs for crystal structure calculations.

5.3.3 Different Ionic Crystals

So far we have considered only NaCl, and other crystals of a similar form. There are various ionic crystal structures, we typically identify these structures by the most common example. The three most common ionic crystal structures are the rock salt type, of which NaCl is an example, caesium chloride type, CsCl, and zinc blende, ZnS.

One scheme for crystal naming is the ***Struktubericht symbols***. In this scheme A refers to crystals of a single atom, B to crystals of two elements in an equal ratio, C to crystals of two elements in a two-to-one ratio, and D) for all other stoichiometry. As well as this there is a number. So A1 is FCC, A2 is BCC, A3 is HCP. However, we are interested in ionic crystals of equal stoichiometry, so B. Rock salt is of type B1, CsCl is of type B2, and ZnS is of type B3. See [Figure 5.4](#).

In all three of these structures both ions have the same local environment. That is, each ion is surrounded by the other ion in the same way such that if you swapped the ions the crystal wouldn't change. Ions of a similar size tend to form the CsCl type whereas ions with a larger size difference tend to form the rock salt type.

The ZnS structure is the same as diamond but if you replaced half of the atoms in diamond with a different type. This structure is common amongst semiconductors used in optoelectronics (such as LEDs), such as GaAs, GaN, and GaP. This type of structure is often found in materials with mixed ionic and covalent bonding.

5.3.4 Ionic Spacing

We can find the equilibrium spacing between ions by finding the value of R which minimises U_{tot} . To do so we need to choose a repulsive potential, we choose $V_r(R) = \lambda e^{-R/\rho}$. We then have

$$0 = \frac{dU_{\text{tot}}}{dR} \Big|_{R=R_0} = N \left[-\frac{z\lambda}{\rho} e^{-R_0/\rho} + \alpha \frac{q^2}{4\pi\epsilon_0 R_0^2} \right]. \quad (5.3.16)$$

Rearranging this we have

$$\left(\frac{R_0}{\rho} \right)^2 e^{-R_0/\rho} = \frac{1}{4\pi\epsilon_0} \frac{\alpha q^2}{\rho z \lambda}. \quad (5.3.17)$$

This has no analytic solution. Numerical solutions give $R_0/\rho \approx 10$. Notice that making either α or q larger will make the crystal denser. The equilibrium energy is

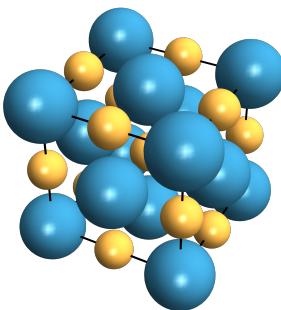
$$U_{\text{tot}}(R_0) = -N\alpha \frac{q^2}{4\pi\epsilon_0 R_0} \left(1 - \frac{\rho}{R_0} \right). \quad (5.3.18)$$

As an example consider KCl. This has the rock salt structure type with $z = 6$. Experimental properties of KCl can be reproduced with $\rho = 0.30 \text{ \AA}$ and $\lambda = 4 \text{ keV}$. We then find $R_0 \approx 10\rho = 3 \text{ \AA}$.

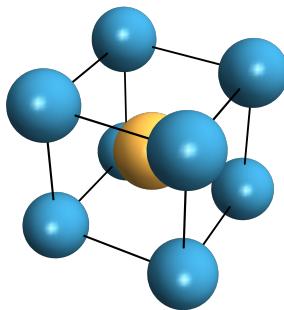
5.4 Covalent Crystals

Covalent bonding arises when the atomic wave functions overlap in such a way as to minimise the energy. Crudely we think of this as “sharing electrons”. Unlike other bonding types this forms directional bonds.

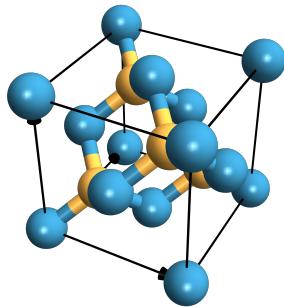
Covalent bonding is more complicated to describe so, rather than give a full quantum mechanical description, we will simply make an analogy to the bonding of the



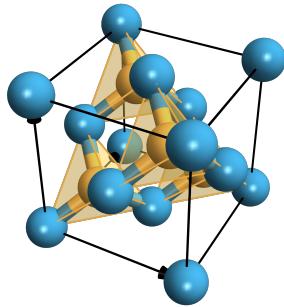
(a) The rock salt (B1) structure.



(b) The caesium chloride (B2) structure.

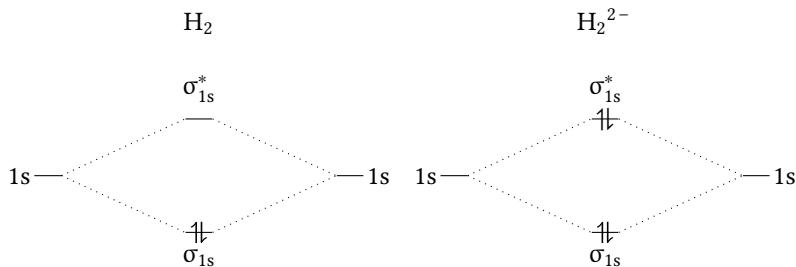


(c) The zinc blende (B3) structure.



(d) The zinc blende (B3) structure.

Figure 5.4: The three most common ionic crystal structures, the rock salt, CsCl, and ZnS types, also known as B1, B2, and B3.

Figure 5.5: Molecular orbital diagrams for H_2 and He_2 .

H_2 molecule². When two hydrogen atoms are brought together their initial atomic, 1s, wave functions are no longer eigenstates. Instead, a new, lower energy, bonding molecular orbital, σ_{1s} , is formed, as well as a higher energy, anti-bonding molecular orbital, σ_{1s}^* . See Figure 5.5. The bonding state has anti-parallel spins and a symmetric wave function, for this reason it is sometimes referred to as an S state. The anti-bonding state has parallel spins and an antisymmetric wave function, for this reason it is sometimes referred to as an A state. For a covalently bonded molecule to be stable the energy must decrease when the wave functions overlap. For this reason H_2 is stable but He_2 is not, since the electrons in the anti-bonding state increase the energy by as much as the electrons in the bonding state decrease it.

²see the notes from principles of quantum mechanics course.

5.5 Metallic Crystals

Metallic bonds are basically more extreme forms of covalent bonds in which the overlap of all wave functions (not just neighbours like in covalent bonds) decrease the energy to form a stable state. The valence electrons become delocalised. The bonds in a metal are non-directional, unlike covalent bonds. We will give a fuller picture later where we treat the delocalised electrons like a Fermi gas.

5.6 Mixed Binding

So far we have only considered one type of bonding at a time. In reality all types of bondings occur at once. Van der Waals bonding is weak enough that most of the time it is negligible compared to the other bonding types. The refractory metals (a subset of the d-block transition metals including Mo and W) have high melting points and hardness since they are a mix of metallic and covalently bonded. Many metal-non-metal crystals mix covalent and ionic bonding in various amounts, ranging from almost pure covalent bonding for Si to about one third ionic bonding for GaAs, to two thirds for ZnS, to almost pure ionic for NaCl.

One interesting case of mixed bonding is graphite. In graphite carbon, which is usually $1s^2 2s^2 2p^4$, forms $1s^2 2s^2 2p^2$ valence electrons and sp^2 hybrid orbitals (covalent bonding) and a delocalised electrons (metallic bonding). This holds together layers of graphite. Between the layers the main bonding effect is van der Waals bonding. This leads to graphites interesting properties like being very conductive along the layers but not across and individual layers being strong but still being soft over all as the layers can easily move past each other.

Six

Phonons

So far we have considered the atoms in a crystal to be stationary at their equilibrium positions, given by minimising the binding energy. We saw, for example, that for helium at low temperature, the vibrational energy of the atoms can effect the physical properties, even to the extent that helium doesn't form a crystal structure at 0 K and standard pressure.

One of the most important quantities that for which we need to take atomic vibrations into account is the heat capacity. We will see three models of heat capacity in increasing complexity and accuracy. Along the way we will introduce the idea of phonons and look at some other consequences.

6.1 Heat Capacity

6.1.1 Classical Heat Capacity

Recall that the heat capacity is defined as the rate of change of internal energy, U , with respect to the temperature, T :

$$C_V := \left(\frac{\partial U}{\partial T} \right)_V. \quad (6.1.1)$$

Strictly the definition above is the heat capacity at constant volume, but, for solids the heat capacity at constant pressure, C_P , is very similar. Since the change in volume as a solid is heated is relatively small compared to a gas. So we usually treat the two as the same. To measure the heat capacity we take a sample of the material, heat it, and measure the corresponding temperature change.

For a monatomic ideal gas the heat capacity is $C_V = 3k_B/2$ per atom, or $C_V = 3k_B N_A/2 = 3R/2$ per mole, where R is the gas constant. In general the gas has a heat capacity of $k_B/2$ per quadratic degree of freedom per atom, and we assume that only translational degrees of freedom are accessible at low temperatures.

For solids the positions of atoms are fixed. Instead the atoms can vibrate about their equilibrium positions. This gives six degrees of freedom, two for each Cartesian direction, one kinetic and one potential. This suggests a heat capacity of $C = 3k_B$ per atom. This is the **Dulong–Petit law**, it suggests a constant, positive heat capacity.

The problem is that experimentally at low temperatures the Dulong–Petit law doesn't hold. In fact we find that $C \rightarrow 0$ as $T \rightarrow 0$.

6.1.2 Einstein Model

¹for more details see the notes on statistical mechanics from the thermal physics course

we treat each atom as being close to its equilibrium position. This allows us to Taylor expand the potential. Since the equilibrium is a minimum the first derivative vanishes and we are left with a quadratic potential, up to an unimportant constant term, given by $V(\Delta r) = C\Delta r^2/2$.



The C here is *not* the heat capacity, it's a force constant. Unfortunately the common alternative of k clashes with the wave vector and so we are left with no good notation for this quantity. Fortunately the force constant doesn't appear in future calculations so this shouldn't be too much of an issue.

We then treat the atoms as quantum harmonic oscillators, in particular we assume they oscillate independently in three dimensions. The energy levels of a one-dimensional harmonic oscillator are

$$E_n = \hbar\omega \left(n + \frac{1}{2} \right) \quad (6.1.2)$$

where $n \in \mathbb{N}$. The average value of n at temperature T is given by the Planck distribution function:

$$\langle n \rangle = \frac{1}{e^{\hbar\omega\beta} - 1} \quad (6.1.3)$$

where $\beta := 1/(k_B T)$. The internal energy is then given by

$$U = 3N\hbar\omega_E \left(\langle n \rangle + \frac{1}{2} \right) \quad (6.1.4)$$

where the factor of 3 comes from considering three-dimensional oscillators. Here $\omega_E = \sqrt{C/m}$ is a parameter that depends on the material with m being the mass of an atom.

The heat capacity can then be found by differentiating:

$$C = \frac{\partial U}{\partial T} \quad (6.1.5)$$

$$= 3N\hbar\omega_E \frac{\partial}{\partial T} \left(\frac{1}{e^{\hbar\omega_E/k_B T} - 1} + \frac{1}{2} \right) \quad (6.1.6)$$

$$= 3Nk_B \left(\frac{\hbar\omega_E}{k_B T} \right)^2 \frac{e^{\hbar\omega_E\beta}}{(e^{\hbar\omega_E\beta} - 1)^2}. \quad (6.1.7)$$

This has only a single parameter, ω_E , which is more commonly given in terms of the **Einstein temperature**, $\Theta_E := \hbar\omega_E/k_B$. This model definitely improves on the Dulong–Petit law. In particular $C \rightarrow 0$ as $T \rightarrow 0$. However, the model still fails to match real data at low temperatures typically giving values of the heat capacity that are too low. This is a fundamental issue with the model due to the assumption of independent oscillators, which is simply not valid, particularly at low temperatures.

For a better model we will need to study vibrations of atoms more carefully, and this is what we do in the rest of this chapter.

6.2 Phonons in One Dimension

Consider a chain of atoms, all of the same type, of mass M , spaced a apart and joined to their neighbours by springs with force constant C . The atoms typically move a

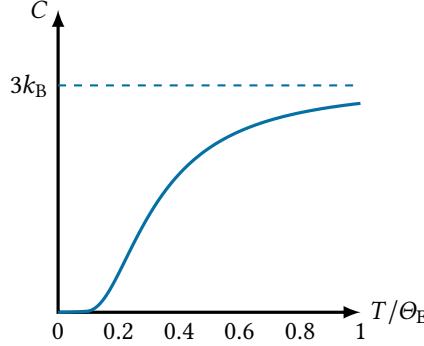


Figure 6.1: The heat capacity as predicted by the Einstein model.

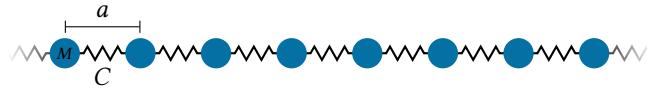


Figure 6.2: A chain of a single type of atom of mass M , separated at equilibrium by a distance a , and connected to their neighbours by a spring with force constant C .

distance 0.05 \AA to 0.1 \AA from their equilibrium positions, that is approximately 2 % of the interatomic separation.

For now we consider only the interactions between nearest neighbours. Let u_s be the displacement of the s th atom. Then the force on this atom is

$$F_s = C(u_{s+1} - u_s) + C(u_{s-1} - u_s) = C(u_{s+1} + u_{s-1} - 2u_s) \quad (6.2.1)$$

where the first term corresponds to the force due to the relative separation of atoms s and $s + 1$ from their equilibrium separation, and the second term corresponds to s and $(s - 1)$'s relative positions. The equation of motion for atom s is

$$M \frac{d^2 u_s}{dt^2} = C(u_{s+1} + u_{s-1} - 2u_s). \quad (6.2.2)$$

We make an ansatz of wavelike solutions:

$$u_s(t) = ue^{i(Ksa - \omega t)} \quad (6.2.3)$$

where K is the wave vector, which relates to the wavelength by $\lambda = 2\pi/K$, u is the amplitude of the wave, and ω is the (angular) frequency of the wave. Atom s is at equilibrium position $x = sa$, taking the zeroth atom's equilibrium position as $x = 0$.

Differentiating the ansatz we have

$$\begin{aligned} u'_s(t) &= -i\omega ue^{i(Ksa - \omega t)} = -i\omega u_s(t), \\ u''_s(t) &= -\omega^2 ue^{i(Ksa - \omega t)} = -\omega^2 u_s(t). \end{aligned}$$

We can also apply the ansatz for u_{s+1} and u_{s-1} getting

$$u_{s+1} = ue^{i(K[s+1]a - \omega t)} = ue^{i(Ksa - \omega t)} e^{iKa} = e^{iKa} u_s, \quad (6.2.4)$$

$$u_{s-1} = ue^{i(K[s-1]a - \omega t)} = ue^{i(Ksa - \omega t)} e^{-iKa} = e^{-iKa} u_s. \quad (6.2.5)$$

Substituting the ansatz into [Equation \(6.2.2\)](#) we get

$$-M\omega^2 u_s = Cu_s(e^{iKa} + e^{-iKa} - 2) = Cu_s[\cos(Ka) - 2]. \quad (6.2.6)$$

From this we get

$$\omega^2 = \frac{2C}{M}[1 - \cos(Ka)] = \frac{4C}{M} \sin^2\left(\frac{1}{2}Ka\right). \quad (6.2.7)$$

Which gives

$$\omega = 2\sqrt{\frac{C}{M}} \left| \sin\left(\frac{1}{2}Ka\right) \right| \quad (6.2.8)$$

This is the **phonon dispersion relation**, it relates the frequency and wave vector. It is periodic in K , with an infinite number of solutions for K for each ω . However, only the wave vectors with $|K| \leq \pi/a$ are physically meaningful. All other wave vectors can be reduced to a wave vector in this range by addition on $2\pi n/a$, for some $n \in \mathbb{Z}$. We call the region of physically meaningful wave vectors the first **Brillouin zone**.

6.3 Wigner–Seitz Cells

The **Wigner–Seitz construction** is a way to define a unique primitive unit cell. The unit cells defined aren't very useful in real space as they cannot be described by three lattice vectors. However, they are still valid unit cells since they tile to fill space in a periodic way with no gaps.

The construction is as follows:

1. Connect a chosen central lattice point to all nearby lattice points.
2. Draw planes perpendicular through the midpoints of all connecting lines.
3. The smallest volume enclosed by these planes is the Wigner–Seitz cell.

This is demonstrated in two dimensions in [Figure 6.3](#). [Figure 6.4](#) then shows how this same cell tiles the plane and contains a single lattice point, making it a valid primitive unit cell.

The main use of Wigner–Seitz cells is to define the Brillouin zone as the Wigner–Seitz cell in reciprocal space. For example the reciprocal lattice of FCC is BCC and the Wigner–Seitz cell of a BCC lattice is the Brillouin zone for the FCC lattice.

6.4 Phonons in Three Dimensions

While treating coupled oscillation in three dimensions is somewhat more complicated the results are fundamentally the same as in one dimension. This is because three-dimensional travelling waves represent movement of whole crystal layers. Instead of a single chain oscillating we have multiple chains oscillating in parallel. If the direction of travel isn't along one of the lattice vectors then we can still decompose the motion into waves along the lattice vectors.

The main difference in three dimensions is that the force constant, C , will usually depend on the direction of the wave vector, K , since crystals are, in general, not isotropic.

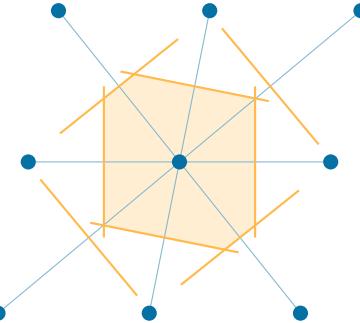


Figure 6.3: The Wigner–Seitz construction of a primitive unit cell.

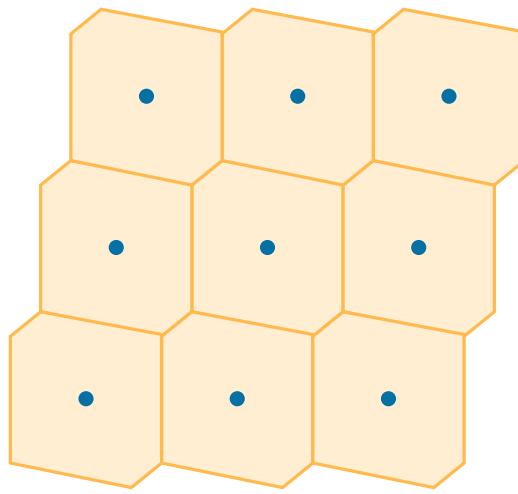


Figure 6.4: A Wigner–Seitz cell is a primitive unit cell, meaning it contains one lattice point and tiles space, as shown here.

It is possible to calculate phonon dispersion relations from first principles from the electronic structure. This is done using a computational framework called **density functional theory** (DFT). It allows for a good approximation to the many-electron quantum mechanical description of the crystal.

When discussing Brillouin zones it is conventional to label lines of symmetry with capital Greek letters, and points of symmetry with capital Latin letters. The one exception is that the Brillouin zone centre, that is $\mathbf{K} = \mathbf{0}$, is always denoted Γ .

To fully plot the dispersion relation would require four dimensions (three to represent \mathbf{k} and one for ω). Instead we typically plot dispersion relations along a line in the Brillouin zone. This is fine since typically we can choose a line such that it goes through all of the interesting points. An example is shown in Figure 6.6. Note that it is conventional to work with frequencies (as opposed to angular frequencies) with phonons, typically phonons range in frequency from 100 Hz to 10 THz.

In three dimensions atoms can oscillate in three orthogonal directions. Along the direction of propagation, giving longitudinal waves, or in along one of two perpendicular directions to the propagation, giving transverse waves.

For a crystal structure with one atom in the primitive unit cell there are three phonon modes per wave vector, however, it is possible that they are degenerate. The

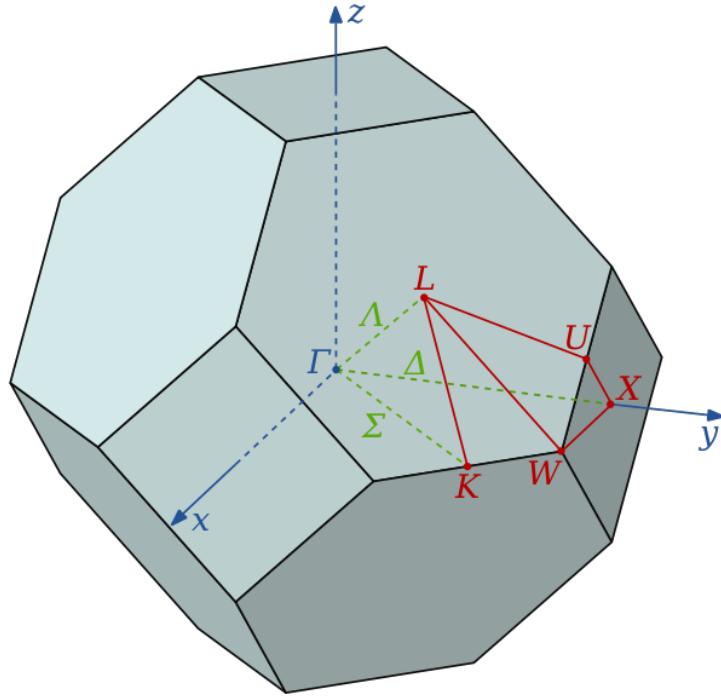


Figure 6.5: The first Brillouin zone for an FCC crystal. The Brillouin zone centre is labelled Γ and the other labelled lines and points are of particularly high symmetry. Taken from [1]

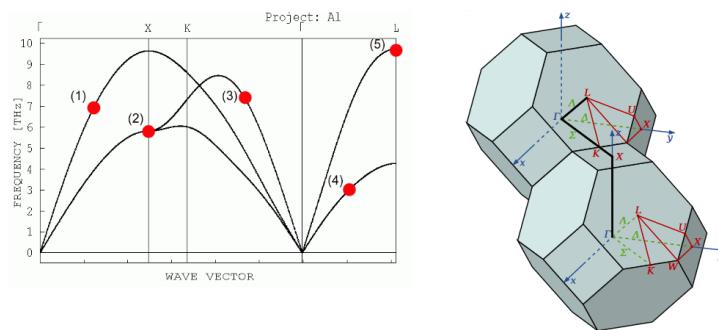


Figure 6.6: Dispersion relation for a sample of FCC aluminium. The dispersion relation is plotted travelling along the black line, starting in the lower Brillouin zone and ending in the equivalent, higher Brillouin zone. Notice that there is degeneracy between Γ and X , as well as between Γ and L , since we can only see two branches at these points. Taken from [2]

longitudinal branch is usually the highest frequency and the transverse branches tend to be degenerate in high symmetry situations.

6.5 Energy Quantisation and Momentum

Consider a quantum harmonic oscillator with frequency ω_0 . The energy levels are given by $E_n = (n + 1/2)\hbar\omega_0$. We can treat the energy of our phonons similarly by substituting ω_0 for the dispersion relation, $\omega(K)$. This gives energy levels

$$E_n(K) = \hbar\omega(K) \left(n + \frac{1}{2} \right) = \hbar\sqrt{\frac{4C}{M}} \left(n + \frac{1}{2} \right). \quad (6.5.1)$$

We can only change the energy in steps of $\hbar\omega(K)$. The total energy must be conserved so if a phonon of energy $\hbar\omega$ is created in some process it must absorb energy $\hbar\omega$ from the process. On the other hand if a phonon of energy $\hbar\omega$ is destroyed in a process then its energy must be absorbed by other parts of the process.

We can treat a phonon as a quantum of vibrational energy, $\hbar\omega(K)$. The phonon also has well defined momentum, $\hbar K$. In this way the phonon is a bit like a particle. The phonon does *not* have well defined position, it's a wave. This is different to particles. We therefore say that phonons are quasiparticles.

The momentum of a phonon is not the same as the momentum in the mechanical sense. The momentum of a given atom will always be offset by the momentum of another particle oscillating exactly out of phase. However, we can still use ideas like momentum conservation.

Consider an elastic scattering of a particle. If it starts with the initial wave vector, \mathbf{k}_i , and ends with the wave vector \mathbf{k}_f , then the crystal must recoil with momentum $-\hbar G$, where G satisfies

$$\mathbf{k}_f = \mathbf{k}_i + G. \quad (6.5.2)$$

G must be a reciprocal lattice vector with integer coefficients. This means that the momentum change is quantised. There are no phonons in this process.

Consider an inelastic scattering of a particle. If a phonon is created in the process then we modify the previous equation to

$$\mathbf{k}_f + \mathbf{K} = \mathbf{k}_i + G, \quad (6.5.3)$$

where \mathbf{K} is the wave vector of the phonon. Similarly if a phonon is destroyed in the process we modify it to

$$\mathbf{k}_f = \mathbf{k}_i + G - \mathbf{K}. \quad (6.5.4)$$

6.6 Measuring Phonon Dispersion

We have only considered nearest neighbours when calculating phonon dispersion relations. Data has been collected that reveals that interactions up to the 10th nearest neighbours can be important enough to give measurable differences in dispersion relations.

Phonon dispersion relations can be measured by measuring the momentum and energy change of scattered particles. Doing so for a variety of angles and initial momenta can allow us to work out possible energy and momentum changes, which can be used to work out the phonon dispersion relation.

6.7 Speed of Sound

Sound waves in solids are simply low frequency phonons, typically with frequency from a few hertz, up to a few million hertz. This is much lower than typically phonon energy which is a few terahertz.

Recall that we have the dispersion relation

$$\omega^2(K) = \frac{2C}{M} [1 - \cos(Ka)]. \quad (6.7.1)$$

The **speed of sound** is defined as the **group velocity**,

$$v_g := \frac{d\omega}{dK}. \quad (6.7.2)$$

Given the low frequency of sound waves the speed of sound corresponds to the slope of the dispersion relation near $K = 0$ (which is the Γ point). We therefore have $Ka \ll 1$, meaning that we can Taylor expand the cos giving $\cos(Ka) \approx 1 - K^2 a^2 / 2$, which then simplifies the dispersion relation to

$$\omega^2(K) \approx \frac{C}{M} K^2 a^2 \implies \omega(K) \approx \sqrt{\frac{C}{M}} a K. \quad (6.7.3)$$

We see that this is linear in K and we can simply read off the derivative as

$$v_g := \frac{d\omega}{dK} \approx a \sqrt{\frac{C}{M}}. \quad (6.7.4)$$

From this we see that the speed of sound, $v_s := v_g \approx \omega/K = \nu\lambda$, where $\nu = 2\pi\omega$ is the frequency and $\lambda = 2\pi/K$ the wavelength.

6.8 Discrete Wave Vectors

Real heat capacity measurements at high temperature give an upper heat capacity of $3R$. Our model does not. The thing we are missing is that real crystals have a finite size. This effects which phonon modes are possible, since the existence of boundaries gives us boundary conditions.

6.8.1 Boundary Conditions in One Dimension

Consider a chain of $N + 1$ atoms with equilibrium separation a , forming a one-dimensional crystal of length $L = Na$. We can have wave solutions with nodes on the surface, i.e. standing waves. These have wavelength $\lambda = 2L/n$. This corresponds to discrete allowed wave vectors

$$K = \frac{\pi n}{L}, \quad \text{where} \quad n = 1, \dots, N - 1. \quad (6.8.1)$$

Now suppose we take the same chain of atoms and bend it into a circle. This shouldn't change the physical properties, we've never observed the macroscopic geometry of a crystal effecting material properties. We can then impose periodic boundary conditions on this circular crystal. For simplicity we make the $(N + 1)$ th atom the first atom, so we have N atoms, and the circumference is $L = Na$. The travelling wave solution is

$$u_s = u(0)e^{i(sKa - \omega_K t)} \quad (6.8.2)$$

where u_s is the displacement of the atom in position s . The phase at any one point must be unique, which means we must have $u_s = u_{s+N}$. This gives the quantisation condition

$$K = \pm \frac{2\pi n}{L}, \quad \text{where} \quad n = 0, \pm 1, \pm 2, \dots, \pm N. \quad (6.8.3)$$

6.8.2 Boundary Conditions in Two Dimensions

In two dimensions the crystal size is its area, which we take to be $A = L_x L_y = N_x N_y a^2$, assuming a rectangular region of crystal with N_x atoms along the x -axis and N_y atoms along the y -axis. Taking the crystal to be periodic in both directions, giving the geometry of a torus, the allowed wave vectors are

$$(K_x, K_y) = (0, 0), \left(\pm \frac{2\pi}{L_x}, \pm \frac{2\pi}{L_y} \right), \dots, \left(\pm \frac{N_x \pi}{L_x}, \pm \frac{N_y \pi}{L_y} \right). \quad (6.8.4)$$

This means that the density of allowed wave vectors is one allowed wave vector per area

$$\frac{2\pi}{N_x a} \frac{2\pi}{N_y a} = \frac{(2\pi)^2}{A} \quad (6.8.5)$$

in reciprocal space.

6.8.3 Boundary Conditions in Three Dimensions

The crystal has volume $V = L_x L_y L_z = N_x N_y N_z a^3$, where we take a cuboidal crystal with edge lengths $L_i = N_i a$, corresponding to N_i atoms. Taking periodic boundary conditions in all three directions, giving the geometry of a hyper-torus, we have allowed wave vectors

$$K_i = 0, \pm \frac{2\pi}{L_i}, \pm \frac{4\pi}{L_i}, \dots, \pm \frac{N_i \pi}{L_i}. \quad (6.8.6)$$

There is one allowed wave vector per

$$\frac{(2\pi)^3}{V} = \frac{8\pi^3}{V}. \quad (6.8.7)$$

6.9 Phonon Density of States

Recall that the heat capacity is defined as

$$C_V := \left(\frac{\partial U}{\partial T} \right)_V. \quad (6.9.1)$$

At a given temperature T there will be a variety of phonon polarisations, p , and wave vectors, \mathbf{K} . The energy of any given phonon with polarisation p and wave vector \mathbf{K} is $\hbar\omega_{\mathbf{K},p}$, where the angular frequency, $\omega_{\mathbf{K},p}$, depends, in general, on both \mathbf{K} and p . The total energy of all of these phonons is then

$$U = \sum_{\mathbf{K}} \sum_p U_{\mathbf{K},p} \quad (6.9.2)$$

where the sums are over all wave vectors and polarisations, and $U_{K,p}$ is the total energy of all phonons with wave vector K and polarisation p . We can replace $U_{K,p}$ with the energy of a single phonon with the same wave vector and polarisation, $\hbar\omega_{K,p}$, scaled by the fraction of phonons that have this energy, which is given by the Planck distribution:

$$\langle n_{K,p} \rangle = \frac{1}{e^{\hbar\omega_{K,p}\beta} - 1}, \quad (6.9.3)$$

where $\beta = 1/(k_B T)$. Hence we have

$$U = \sum_K \sum_p U_{K,p} \quad (6.9.4)$$

$$= \sum_K \sum_p \langle n_{K,p} \rangle \hbar K, p \quad (6.9.5)$$

$$= \sum_K \sum_p \frac{\hbar\omega_{K,p}}{e^{\hbar\omega_{K,p}\beta} - 1}. \quad (6.9.6)$$

Importantly we only use K and p as labels for the phonons, what actually matters are the energies, $\hbar\omega_{K,p}$, and the distribution of the energy. In particular the number of phonons with energies in the range $\hbar\omega$ to $\hbar(\omega + \Delta\omega)$ for some small $\Delta\omega$. For this reason we use the density of states, $D(\omega)$, which is such that $dN = D(\omega) d\omega$, where dN is the change in the number of phonons associated with the frequency change $\omega \rightarrow \omega + d\omega$.

Integrating the density of states we have

$$N = \int dN = \int D(\omega) d\omega. \quad (6.9.7)$$

So this gives the number of states in a given range. We can find the density of states by rearranging the definition to give

$$D(\omega) = \frac{dN}{d\omega}. \quad (6.9.8)$$

For a given phonon dispersion relation, which we assume is isotropic so $\omega(K) = \omega(K)$, the density of states is

$$D(\omega) = \frac{dN}{d\omega} = \frac{dN}{dK} \frac{dK}{d\omega}. \quad (6.9.9)$$

We can interpret dN/dK as the density of states in reciprocal space, which can be found by considering the geometry of the space, and we can find the value of $dK/d\omega = (d\omega/dK)^{-1}$ from the dispersion relation.

6.9.1 Density of States in One Dimension

In one dimension there is one allowed wave vector per $2\pi/L$, hence dN/dK , the number density of allowed states, is just the reciprocal of this:

$$\frac{dN}{dK} = \left(\frac{2\pi}{L} \right)^{-1} = \frac{L}{2\pi}. \quad (6.9.10)$$

We get $dK/d\omega$ from $1/(\frac{d\omega}{dK})$, which is the gradient of the dispersion relation. We therefore get a singularity when the dispersion relation is horizontal, which happens at the edge of the first Brillouin zone.

6.9.2 Density of States in Three Dimensions

In three dimensions there is one allowed wave vector per $8\pi^3/V_{\text{crystal}}$. Consider the a sphere in reciprocal space of radius K . The volume of this sphere is

$$V_{\text{sphere}}(K) = \frac{4}{3}\pi K^3. \quad (6.9.11)$$

The number of states in this sphere is then

$$N_{\text{sphere}}(K) = \frac{4}{3}\pi K^3 \left(\frac{8\pi^3}{V_{\text{crystal}}}\right)^{-1} = \frac{K^3 V_{\text{crystal}}}{6\pi^2}. \quad (6.9.12)$$

It follows that

$$\frac{dN_{\text{sphere}}}{dK} = \frac{K^2 V_{\text{crystal}}}{2\pi^2}. \quad (6.9.13)$$

The density of states is then

$$D(\omega) = \frac{K^2 V_{\text{crystal}}}{2\pi^2} \frac{dK}{d\omega}. \quad (6.9.14)$$

We can do this same calculation in a slightly different way by considering spherical shells of radius K and thickness dK . The volume of such a shell is

$$dV_{\text{shell}}(K) = 4\pi K^2 dK, \quad (6.9.15)$$

which follows from

$$dV = \frac{4}{3}\pi^2 [(K + dK)^3 - K^3] \quad (6.9.16)$$

$$= \frac{4}{3}\pi^2 [K^3 + 3K^2 dK + 3K dK^2 + dK^3 - K^3] \quad (6.9.17)$$

$$= 4\pi K^2 dK + O(dK^2). \quad (6.9.18)$$

The number of states in the shell is

$$dN_{\text{shell}}(K) = 4\pi K^2 dK \left(\frac{8\pi^3}{V_{\text{crystal}}}\right)^{-1} + O(dK^2) = \frac{K^2 V_{\text{crystal}}}{2\pi^2} dK + O(dK^2). \quad (6.9.19)$$

Hence the density of states is

$$\frac{dN_{\text{shell}}}{dK} = \frac{K^2 V_{\text{crystal}}}{2\pi^2} + O(dK), \quad (6.9.20)$$

which is the same result as before in the limit $dK \rightarrow 0$.

In three dimensions there are significantly more stationary points in the dispersion relation, and hence poles in the density of states. The density of states quickly gets quite complicated, but treating such cases is beyond this course.

6.10 Heat Capacity

The heat capacity at constant volume is defined as

$$C_V := \left(\frac{\partial U}{\partial T}\right)_V. \quad (6.10.1)$$

The total energy of all phonons in the system, at some fixed temperature, T , is characterised by the polarisation, p , and wave vector, \mathbf{K} , which give us the phonon energy, $\hbar\omega_{\mathbf{K},p}$. We saw in [Equation \(6.9.6\)](#) that the total energy stored in the lattice, that is the energy of the phonons, is

$$U_{\text{lat}} = \sum_{\mathbf{K}} \sum_p \frac{\hbar\omega_{\mathbf{K},p}}{e^{\hbar\omega_{\mathbf{K},p}\beta} - 1}. \quad (6.10.2)$$

Using the density of states, $D_p(\omega)$, of phonons of a given polarisation, p , we find that

$$U_{\text{lat}} = \sum_p \int_0^\infty D_p(\omega) \frac{\hbar\omega}{e^{\hbar\omega\beta} - 1} d\omega. \quad (6.10.3)$$

Now we have

$$\frac{\partial}{\partial T} = \frac{\partial\beta}{\partial T} \frac{\partial}{\partial\beta} = -\frac{1}{k_B T^2} \frac{\partial}{\partial\beta} \quad (6.10.4)$$

and so the lattice heat capacity is

$$C_{\text{lat}} = -\frac{1}{k_B T^2} \frac{\partial}{\partial\beta} \sum_p \int_0^\infty D_p(\omega) \frac{\hbar\omega}{e^{\hbar\omega\beta} - 1} d\omega \quad (6.10.5)$$

$$= \frac{1}{k_B T^2} \sum_p \int_0^\infty D_p(\omega) \frac{(\hbar\omega)^2 e^{\hbar\omega\beta}}{(e^{\hbar\omega\beta} - 1)^2} d\omega \quad (6.10.6)$$

$$= k_B \sum_p \int_0^\infty \int_0^\infty \frac{x^2 e^x}{(e^x - 1)^2} d\omega \quad (6.10.7)$$

where $x = \hbar\omega\beta$.

6.10.1 Einstein Model

In the Einstein model we have $3N$ identical, uncoupled harmonic oscillators, all of frequency ω_0 . This is equivalent to a density of states

$$D(\omega) = 3N\delta(\omega - \omega_0), \quad (6.10.8)$$

and a single phonon branch. The heat capacity is then

$$C_{\text{lat}} = k_B \int_0^\infty 3N\delta(\omega - \omega_0) \frac{x^2 e^x}{(e^x - 1)^2} d\omega \quad (6.10.9)$$

$$= 3Nk_B (\hbar\omega_0\beta)^2 \frac{e^{\hbar\omega_0\beta}}{(e^{\hbar\omega_0\beta} - 1)^2}. \quad (6.10.10)$$

As previously discussed this tends to the correct constant heat capacity at high temperatures and tends to zero as $T \rightarrow 0$, but experimental data shows that at low temperatures the heat capacity should go as T^3 , which this doesn't.

6.10.2 Debye Model

The **Debye model** manages to correctly encapsulate the T^3 behaviour of the heat capacity at low temperatures, while still depending on only a single material parameter. It does so without us having to know the full phonon density of states.

At low temperatures most phonon states are unattainable. The only states that can be occupied are the low energy, acoustic frequency, states. Here the dispersion relation is approximately linear. We imagine N oscillators in a crystal of volume V and a simplified dispersion relation, $\omega(K) = vK$ where v is the sound velocity, since these low wave number phonons are exactly the phonons that form sound waves. Hence, $K = \omega/v$. The density of states, in three dimensions, is then simply

$$D(\omega) = \frac{VK^2}{2\pi^2} \frac{\partial K}{\partial \omega} = \frac{V\omega^2}{2\pi^2 v^3}. \quad (6.10.11)$$

One problem that we face is this doesn't correctly give us N when we integrate over all frequencies. To fix this we introduce a cut-off frequency, called the **Debye frequency**, ω_D , which is such that

$$N = \int_0^{\omega_D} D(\omega) d\omega. \quad (6.10.12)$$

This is a reasonable thing to do since in this model higher frequency states are largely unoccupied. Computing this integral we get

$$N = \int_0^{\omega_D} \frac{V\omega^2}{2\pi^2 v^3} d\omega = \frac{V\omega_D^3}{6\pi^2 v^3} \quad (6.10.13)$$

from which we get

$$\omega_D = \left(6\pi^2 v^3 \frac{N}{V} \right)^{1/3}, \quad (6.10.14)$$

which is useful since $n = N/V$ is the number density, which is relatively easy to work out for a given material, we can simply consider the number of atoms in a unit cell and the volume of the unit cell.

We can define the related quantity, the **Debye wave vector**, as

$$K_D := \frac{\omega_D}{v} = \left(6\pi^2 \frac{N}{V} \right)^{1/3}. \quad (6.10.15)$$

We are considering a three dimensional crystal, which will have three acoustic branches. Hence the phonon internal energy includes a factor of three and is given by

$$U = 3 \int_0^{\omega_D} D(\omega) \hbar\omega \langle n(\omega) \rangle d\omega \quad (6.10.16)$$

where $\langle n(\omega) \rangle$ is the Planck distribution, which means that $\hbar\omega \langle n(\omega) \rangle$ is the average energy of an oscillator. Inserting the Debye density of states and the Planck distribution we have

$$U = 3 \int_0^{\omega_D} \frac{V\omega^2}{2\pi^2 v^3} \frac{\hbar\omega}{e^{\hbar\omega\beta} - 1} d\omega \quad (6.10.17)$$

$$= \frac{3V\hbar}{2\pi^2 v^3} \int_0^{\omega_D} \frac{\omega^3}{e^{\hbar\omega\beta} - 1} d\omega. \quad (6.10.18)$$

We can express ω_D in terms of the sound velocity, v , and oscillator density, $n = N/V$, which means that this form of the internal energy depends only on these two material parameters. We can calculate the heat capacity by computing the derivative:

$$C_V = \left(\frac{\partial U}{\partial T} \right)_V \quad (6.10.19)$$

$$= -\frac{1}{k_B T^2} \left(\frac{\partial U}{\partial \beta} \right)_V \quad (6.10.20)$$

$$= \frac{1}{k_B T^2} \frac{3V\hbar}{2\pi^2 v^3} \int_0^{\omega_D} \frac{\hbar\omega^4 e^{\hbar\omega\beta}}{(e^{\hbar\omega\beta} - 1)^2} d\omega. \quad (6.10.21)$$

Introducing the dimensionless parameter $x = \hbar\omega\beta$, and the **Debye temperature** $\vartheta_D = \hbar\omega_D/k_B$ it can be shown that

$$C_V = 9Nk_B \left(\frac{T}{\vartheta_D} \right)^3 \int_0^{\vartheta_D/T} \frac{x^4 e^x}{(e^x - 1)^2} dx. \quad (6.10.22)$$

The prefactor of N shows that the heat capacity depends on the number of oscillators, that is the amount of material. This is a statement that the heat capacity is an extrinsic material, which we know, simply put “more stuff, more ability to store heat”. The only material parameter here is the Debye temperature, ϑ_D , which relates to the speed of sound and oscillator density by

$$\vartheta_D = v \frac{\hbar}{k_B} \left(6\pi^2 \frac{N}{V} \right)^{1/3}. \quad (6.10.23)$$

The Debye temperature turns out to be a very useful material parameter, the heat capacity is related to the energy stored in phonons, which in turn depends on the strength of bonds, which relates to many other properties, such as the strength of the material.

The most important feature of this equation for the heat capacity is its low temperature limit. It's not easy to figure out directly from this point though, instead we return to the internal energy, which we write as

$$U = \frac{3V}{2\pi^2 v^3} \frac{k_B^4 T^4}{\hbar^3} \int_0^{x_D} \frac{x^3}{e^x - 1} dx \quad (6.10.24)$$

where $x_D = \hbar\omega_D\beta$. We then can write this as

$$U = 9Nk_B T \left(\frac{T}{\vartheta_D} \right)^3 \int_0^{\vartheta_D/T} \frac{x^3}{e^x - 1} dx. \quad (6.10.25)$$

In the limit of $T \rightarrow \infty$ the upper limit of the integral goes to infinity. We then have my favourite integral, which it can be shown is

$$\int_0^\infty \frac{x^3}{e^x - 1} dx = \frac{\pi^4}{15}. \quad (6.10.26)$$

See [Chapter B](#) for details. We then have

$$C_V \approx \frac{12}{5} \pi^4 N k_B \left(\frac{T}{\vartheta_D} \right)^3 \quad (6.10.27)$$

which is proportional to T^3 , which agrees with experiments for most materials at low temperatures.

6.10.3 Beyond the Debye Model

The Einstein model has a delta distribution as a density of states. The Debye model improves on this with a density of states proportional to ω^2 , and a cut-off frequency ω_D . There are many more approximations we can make (or not make) to get more accurate densities of states. They quickly become very complicated, and tend to be material specific. We'll stop with the Debye model as it is accurate enough for many purposes and still very general.

6.11 Diatomic Linear Chain

Many materials have a crystal structure with more than one atom per unit cell. We take as a model for such materials a chain of two alternating atoms with different masses, M_1 and M_2 , forming a unit cell of length a . This is the one dimensional analogue of materials like NaCl. We consider only the interactions with nearest neighbours, which we take to form a harmonic potential with force constant C .

Nearest neighbours are always of alternating types, so we only need to consider one force constant for interactions between the two different types of atoms. The distance between nearest neighbours is $a/2$, since the cell now repeats every two atoms. We index the atoms such that u_s is the position of one atom and v_s is the position of the next atom of the opposite type, and then u_{s+1} is the position of the next atom, back to the first type, and so on. We get two equations of motion:

$$M_1 \frac{d^2 u_s}{dt^2} = C(v_s + v_{s-1} - 2u_s), \quad (6.11.1)$$

$$M_2 \frac{d^2 v_s}{dt^2} = C(u_{s+1} + u_s - 2v_s). \quad (6.11.2)$$

Making the ansatz of travelling wave solutions,

$$u_s = ue^{isKa}e^{-i\omega t}, \quad \text{and} \quad v_s = ve^{isKa}e^{-i\omega t}, \quad (6.11.3)$$

the first equation gives

$$-M_1 \omega^2 ue^{sKa}e^{-i\omega t} = C(ve^{sKa} + ve^{(s-1)Ka} - 2ue^{sKa})e^{-i\omega t} \quad (6.11.4)$$

$$\implies -M_1 \omega^2 = Cv(1 + e^{-K_a}) - 2u, \quad (6.11.5)$$

and similarly, for the second equation, we get

$$-\omega^2 M_2 v = Cu(e^{iKa} + 1) - 2Cv. \quad (6.11.6)$$

Given a set of homogeneous linear equations,

$$au + bv = 0, \quad \text{and} \quad cu + dv = 0, \quad (6.11.7)$$

there is a non-trivial solution if and only if

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc = 0. \quad (6.11.8)$$

In our case this corresponds to the requirement that, for a non-trivial solution, we must have

$$\begin{vmatrix} 2C - M_1 \omega^2 & -C(1 + e^{-iKa}) \\ -C(1 + e^{iKa}) & 2C - M_2 \omega^2 \end{vmatrix} = 0 \quad (6.11.9)$$

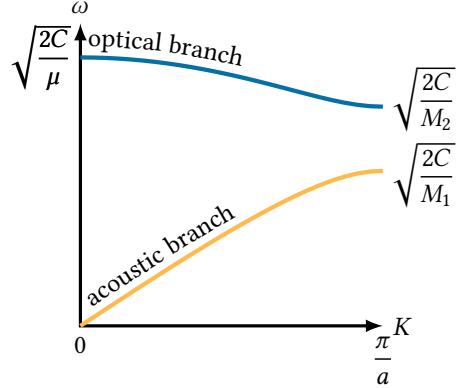


Figure 6.7: The dispersion relation for a diatomic linear chain. Note that there are two branches, optical and acoustic. $\mu := M_1 M_2 / (M_1 + M_2)$ is the reduced mass. As plotted $M_1 = 2M_2$.

from which we get

$$M_1 M_2 \omega^4 - 2C(M_1 + M_2)\omega^2 + 2C^2(1 - \cos(Ka)) = 0. \quad (6.11.10)$$

This is quadratic in ω^2 with solutions

$$\omega^2 = \frac{2C(M_1 + M_2) \pm \sqrt{4C^2(M_1 + M_2)^2 - 8M_1 M_2 C^2(1 - \cos(Ka))}}{2M_1 M_2}. \quad (6.11.11)$$

This is plotted in Figure 6.7. There are two positive solutions, which we refer to as branches. We call the higher frequency one the optical branch, as we will see these phonons are easily excited by visible light. We call the lower frequency the acoustic branch, as these phonons are responsible for sound waves.

6.11.1 Zone-Centre Solutions

Defining the reduced mass,

$$\mu := \frac{M_1 M_2}{M_1 + M_2}, \quad (6.11.12)$$

we see that for $K = 0$ we have $\cos(kA) = 1$, and so for the optical branch we have

$$\omega_+ = \sqrt{\frac{2C(M_1 + M_2)}{M_1 M_2}} = \sqrt{\frac{2C}{\mu}}. \quad (6.11.13)$$

For the acoustic branch we have $\omega_-(0) = 0$. We can also Taylor expand to get

$\cos(Ka) \approx 1 - K^2 a^2 / 2$, so

$$\omega_-^2 \approx \frac{2C(M_1 + M_2) - \sqrt{4C^2(M_1 + M_2)^2 - 8M_1 M_2 C^2(K^2 a^2 / 2)}}{2M_1 M_2} \quad (6.11.14)$$

$$= \frac{2C(M_1 + M_2) \left[1 - \sqrt{1 - \frac{8M_1 M_2 C^2}{4C^2(M_1 + M_2)^2} K^2 a^2 / 2} \right]}{2M_1 M_2} \quad (6.11.15)$$

$$\approx \frac{2C(M_1 + M_2) \left[1 - 1 + \frac{1}{2} \frac{8M_1 M_2 C^2}{4C^2(M_1 + M_2)^2 K^2 a^2 / 2} \right]}{2M_1 M_2} \quad (6.11.16)$$

$$= \frac{C/2}{M_1 + M_2} K^2 a^2. \quad (6.11.17)$$

Which gives

$$\omega_- \approx \sqrt{\frac{C/2}{M_1 + M_2}} aK = v_s K \quad (6.11.18)$$

where we define

$$v_s := \sqrt{\frac{C/2}{M_1 + M_2}}. \quad (6.11.19)$$

We can then identify this quantity as

$$v_s \approx \frac{d\omega}{dK}, \quad (6.11.20)$$

which is to say that v_s is the group velocity of phonons in the acoustic branch with low wave number, identifying phonons in this range as having frequencies in the range of audible sound we see that v_s is the speed of sound in the material.

6.11.2 Zone-Boundary Solutions

On the other side of the Brillouin zone we have $K = \pm\pi/a$, and so $\cos(Ka) = -1$. We then have

$$\omega^2 = \frac{2C(M_1 + M_2) \pm \sqrt{4C^2(M_1 + M_2)^2 - 16M_1 M_2 C^2}}{2M_1 M_2} \quad (6.11.21)$$

$$= \frac{2C(M_1 + M_2) \pm 2C\sqrt{(M_1 + M_2)^2 - 4M_1 M_2}}{2M_1 M_2} \quad (6.11.22)$$

$$= \frac{C(M_1 + M_2 \pm C(M_1 - M_2))}{M_1 M_2} \quad (6.11.23)$$

where we have identified

$$(M_1 + M_2)^2 - 4M_1 M_2 = M_1^2 + M_2^2 - 2M_1 M_2 = (M_1 - M_2)^2. \quad (6.11.24)$$

We therefore have

$$\omega_+ = \sqrt{\frac{2C}{M_2}}, \quad \text{and} \quad \omega_- = \sqrt{\frac{2C}{M_1}}, \quad (6.11.25)$$

for the optical and acoustic branch respectively.

The two zone boundary frequencies each depend only on the mass of *one* of the types of particle. This suggests that for a given phonon on the zone boundary only one type of atom is moving, which type depends on which branch the phonon is on.

6.11.3 Acoustic and Optical Phonons

We now come to the names, acoustic and optical. In the acoustic branch neighbouring atoms move in the same direction. Sound waves in the long wavelength limit also have this property, hence we call these acoustic phonons.

In the optical branch neighbouring atoms move in opposite directions. If the two types of atoms have opposite charges, as is, for example, the case in NaCl, then we essentially have oscillating dipoles, which can couple to electromagnetic radiation, hence we call these optical phonons.

6.11.4 Polarisation

Phonons can be polarised into one of two types. First, longitudinal vibrations, in which the motion of the atoms is parallel to the direction of propagation. Second, transverse vibrations, in which the motion of the atoms is perpendicular to the direction of propagation.

In three dimensions one phonon is longitudinal and two are transverse.

6.11.5 Multi-Atom Phonons in 3D

In three dimensions a three-dimensional primitive unit cell with p atoms has $3p$ oscillator modes. There will be three acoustic branches and the $3p - 3$ other branches will be optical. Including polarisation there are four possible types of phonon, transverse acoustic, longitudinal acoustic, transverse optical, and longitudinal optical.

■ Example 6.11.26 — Diamond Phonons Diamond has two atoms per primitive unit cell. Hence it has 6 phonon branches. Of these 3 are acoustic, 2 of which are transverse acoustic and the other is longitudinal acoustic. The other 3 are optical, of which 2 are transverse optical and the final branch is optical longitudinal.

6.12 Thermal Expansion

Most materials expand when heated. The standard explanation is that atoms in the heated material vibrate more and hence take up more room. This simplified explanation isn't wrong exactly, but the reality isn't quite this simple. While the amplitude of oscillation does increase in our model so far since we assume a harmonic potential the vibrations are equally likely to be going in any direction, and hence the average bond length remains constant and the material doesn't expand.

To mathematically describe the expansion we need an anharmonic potential. We will do so for a single chain of atoms of mass M , spaced a apart. We consider the anharmonic potential

$$U(x) = cx^2 - gx^3 - fx^4 \quad (6.12.1)$$

for some positive constants c , g , and f . Note that x is the displacement from equilibrium, not the position. The $-gx^3$ term provides the asymmetry needed for bond length to not remain the same on average, but instead preferentially increase, and the $-fx^4$ term corresponds to bond softening at large x . We assume that g and f are small, so that the corrections are small compared to the harmonic potential, that is $|gx^3|fx^4 \ll cx^2$.

We consider the average displacement, $\langle x \rangle$, as a function of the temperature. We assume classical statistics, and hence the average displacement is given by the Boltzmann distribution:

$$\langle x \rangle = \frac{\int_{-\infty}^{\infty} x e^{-U(x)\beta} dx}{\int_{-\infty}^{\infty} e^{-U(x)\beta} dx}. \quad (6.12.2)$$

We can evaluate this by expanding the exponential, $e^z = 1+z+z^2/2+\dots$, in particular for the numerator we have

$$x e^{-\beta(cx^2-gx^3-fx^4)} = x e^{-\beta cx^2} e^{\beta g x^3} e^{\beta f x^4} \quad (6.12.3)$$

$$= x e^{-\beta cx^2} [1 + \beta g x^3 + O(x^6)] [1 + \beta f x^4 + O(x^8)] \quad (6.12.4)$$

$$= e^{-\beta cx^2} [x + \beta g x^4 + \beta f x^5 + O(x^7)]. \quad (6.12.5)$$

The denominator only provides the overall scaling normalisation factor. Therefore it is less important and we keep only the dominant quadratic term:

$$e^{-\beta U(x)} \approx e^{-\beta cx^2}. \quad (6.12.6)$$

We then have

$$\langle x \rangle \approx \frac{\int_{-\infty}^{\infty} e^{-\beta cx^2} [x + \beta g x^4 + \beta f x^5] dx}{\int_{-\infty}^{\infty} e^{-\beta cx^2} dx}. \quad (6.12.7)$$

Now, $e^{-\beta cx^2}$ is an even function, so the x and x^5 terms in the numerator vanish in the integral since they give an overall odd integrand. We are then left with

$$\langle x \rangle \approx \frac{\int_{-\infty}^{\infty} \beta g x^4 e^{-\beta cx^2} dx}{\int_{-\infty}^{\infty} e^{-\beta cx^2} dx}. \quad (6.12.8)$$

We can then use standard integrals, or a computer algebra system, to evaluate these. We get

$$\int_{-\infty}^{\infty} \beta g x^4 e^{-\beta cx^2} dx = \frac{3g\sqrt{\pi}}{4c^{5/2}\beta^{3/2}} \quad (6.12.9)$$

and

$$\int_{-\infty}^{\infty} e^{-\beta cx^2} dx = \sqrt{\frac{\pi}{c\beta}}. \quad (6.12.10)$$

■ Code 6.12.11

```

1 Integrate[
2   β g x^4 Exp[-β c x^2],
3   {-x, -∞, ∞},
4   Assumptions → {β > 0, c > 0, g > 0}
5   ] // PowerExpand // Simplify

```

```

6  $\frac{3g\sqrt{\pi}}{4c^{5/2}\beta^{3/2}}$ 
7
8 Integrate[
9   Exp[- $\beta c x^2$ ],
10  {x, - $\infty$ ,  $\infty$ },
11  Assumptions -> { $\beta > 0$ ,  $c > 0$ }
12 ]
13  $\frac{\sqrt{\pi}}{\sqrt{c\beta}}$ 

```

Combining this we have

$$\langle x \rangle \approx \frac{3g\sqrt{\pi}/(4c^{5/2}\beta^{3/2})}{\sqrt{\pi}/(c^{1/2}\beta^{1/2})} = \frac{3g}{4c^2} \frac{1}{\beta} = \frac{3g}{4c^2} k_B T. \quad (6.12.12)$$

This is the average change in bond length with temperature, from the length at $T = 0$ K.

A result of this is that the lattice parameter, as a function of temperature, $a(T)$, is approximately linear at high temperatures, which is the region in which our assumptions are valid. What exactly counts as high temperatures is material dependent but we can assume θ_D or larger. The slope of $a(T)$ gives the **coefficient of linear thermal expansion**, which is defined as

$$\alpha = \frac{1}{L} \frac{dL}{dT} = \frac{1}{a} \frac{da}{dT}, \quad (6.12.13)$$

where L is the length of the crystal, which is proportional to the length of the unit cell. The factor of $1/L$ essentially normalises for the number of unit cells, and so samples of different sizes should have the same value of α .

The model breaks down around room temperature, when we start to see non-linear changes in a . The model breaks further at low temperatures, when we can have weird properties like α being negative, which comes from select phonon modes. Regardless, our model works fairly well for most simple crystals at reasonable temperatures.

6.13 Lattice Thermal Conductivity

When we heat one side of an object the other side also gets hot. Somehow the energy is crossing the object. Here we consider one mechanism involved in this process, namely thermal conductivity due to phonons. This is sufficient to describe thermal conduction in insulators. In metals the free electrons also contribute to the thermal conductivity, which we shall account for later.

We consider heat transfer along a rod, which has temperature $T(x)$ at position x . Heat conduction is a diffusive process, and as such we expect it to follow Fick's first law, which states that

$$J = -D \nabla n \quad (6.13.1)$$

where J is the flux of the relevant quantity, D is the diffusion coefficient, and n is the density of the relevant property. In analogy to this we look for an equation of the form

$$j_U = -\kappa \frac{dT}{dx} \quad (6.13.2)$$

where j_U is the flux of thermal energy, which is the energy per unit area per unit time, and κ is the thermal conductivity coefficient.

For diffusion in a gas we have $D = v_t \ell \cdot v_t$ where v_t is the mean thermal velocity and ℓ is the mean free path. In analogy to this we assume that κ is of the form

$$\kappa = \frac{1}{3} C v \ell \quad (6.13.3)$$

where C is the phonon heat capacity, v the average phonon velocity, and ℓ the phonon mean free path. The heat capacity is added in as we are interested in the energy that transfers through the system, rather than the number of particles, so we need to account for how much energy each particle carries.

Without scattering the mean free path, and hence thermal conductivity, would be infinite. Clearly this isn't the case, thermal conduction takes time, and so we need to consider phonon scattering. This is the topic of the next section.

6.13.1 Phonon Scattering

There are two main types of scattering which limit the mean free path. The first is **geometrical scattering**, which is scattering due to crystal imperfections, such as grain boundaries and crystal surfaces. The second is phonon-phonon scattering. In an ideal crystal with a purely harmonic potential the phonon modes are all independent and don't interact, so there is no phonon-phonon scattering. Therefore we must consider an anharmonic contribution to the potential. Phonon-phonon scattering occurs even in a perfect crystal, and is usually the dominant cause of phonon scattering at room temperature and above.

Geometrical phonon scattering limits the mean free path and hence thermal conductivity. However, it is an elastic process, the energy of the phonon and the magnitude of the wave vector are unchanged, only the direction changes. This means that geometrical scattering cannot redistribute energy between phonon modes, which is required to reach a local equilibrium.

Phonon-phonon scattering must be inelastic, otherwise there would be no way to reach equilibrium. In phonon-phonon scattering two phonons interact and combine to form a new phonon. The overall phonon energy wave vector are conserved in this process, but the energy and wave vector are redistributed among the three phonons involved (the two interacting phonons and the resulting phonon). There is also a reverse process where a phonon decays into two new phonons. There are also scattering processes involving three or more phonons, although these are suppressed by the need for three or more phonons to combine at once, which is fairly unlikely.

We consider two types of phonon-phonon scattering. In **normal scattering**, or N-scattering, two phonons with wave vectors K_1 and K_2 combine to form a phonon with wave vector $K_3 = K_1 + K_2$. This is what we would expect from normal scattering events involving real particles.

In **umklapp scattering**², or U-scattering, two phonons with wave vectors K_1 and K_2 combine, however, if they produced a wave vector $K_1 + K_2$ then this would

²from the German *umklappen* for "to flip over"

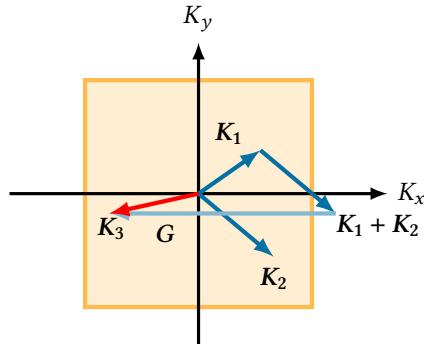


Figure 6.8: Umklapp scattering. The sum of the wave vectors, $K_1 + K_2$, is outside of the Brillouin zone and so the resulting phonon has its wave vector shifted by G so that it falls in the Brillouin zone. In the process the direction of the wave vectors is reversed from pointing in the positive K_x direction to the negative K_x direction.

be outside of the Brillouin zone. Instead the resulting phonon has the wave vector K_3 satisfying

$$K_1 + K_2 = K_3 + G \quad (6.13.4)$$

where G is a reciprocal lattice vector such that K_3 is in the first Brillouin zone.

In umklapp scattering we require that the wave vectors of the two phonons are large enough that the resulting wave vector leaves the first Brillouin zone. This requirement corresponds to them having large energies, which is unlikely at low temperatures. For this reason the type of scattering that dominates depends on the temperature, with normal scattering being more important near $T = 0\text{ K}$, and umklapp scattering becoming more important as temperature increases.

In umklapp scattering if the sum of the wave vectors of the initial phonons is in the first half of the neighbouring Brillouin zone then once mapped back into the Brillouin zone the result will be in the far side of the Brillouin zone and tend to point in the opposite direction to the initial phonons, so it is as if the phonons combine and start travelling backwards. This is demonstrated in Figure 6.8.

6.13.2 Temperature Dependence

The phonon thermal conductivity depends on temperature in a fairly complex way. Broadly we split into three regions which are each dominated by a different behaviour.

At high temperatures, $T > \vartheta_D$, which corresponds to $k_B T > \hbar\omega_D$, all phonon modes are excited, including those with large values of K . The umklapp phonon scattering process is the most important. The number of photons is proportional to the temperature, and the mean free path is inversely proportional to the number of phonons, so

$$\kappa \propto \ell \propto \frac{1}{n} \propto \frac{1}{T}. \quad (6.13.5)$$

So the thermal conductivity increases with decreasing temperature.

At low temperatures, where $T \lesssim \vartheta_D$, there is a smaller fraction of high energy phonons, where we define ‘‘high energy’’ to mean $\hbar\omega > k_B\vartheta_D/2$, which is approximately the average energy required in a scattering event for the resulting sum of

wave vectors to be outside the Brillouin zone. This means that umklapp scattering is less likely. The number of suitable phonons decreases as the Boltzmann factor

$$\exp \left[-\frac{\frac{1}{2}k_B\vartheta_D}{k_B T} \right]. \quad (6.13.6)$$

Hence

$$\kappa \propto e^{\vartheta_D/(2T)} \quad (6.13.7)$$

and so thermal conductivity increases with decreasing temperature.

In the very low temperature regime, $T \ll \vartheta_D$, the mean free path is approximately given by the distance between crystal defects, since phonon-phonon scattering becomes minimal. This value is constant and so

$$\kappa = \frac{1}{3} C v \ell \propto T^3 \quad (6.13.8)$$

since $C \propto T^3$, and v and ℓ are approximately constant. The thermal conductivity then decreases with decreasing temperature.

Seven

Free Electron Gas

To exactly model a solid we would need to contend with approximately 10^{23} atoms per centimetre cubed. All nuclei and electrons interact with all other nuclei and electrons via both electromagnetic effects and quantum effects, such as the Pauli exclusion principle. Clearly this is completely intractable and we need drastic simplifications to be able to get a workable model.

One set of approximations is the **Born–Oppenheimer approximation**, also known as the **adiabatic approximation**. In this we decouple the motion of ionic cores and electrons, essentially assuming that the heavy, relatively slow moving, ionic cores are stationary, and simply act to keep the electrons inside the material.

This is useful since many properties relate mostly to electronic *or* atomic motion, allowing us to ignore one component of the model. For example, the phonon vibrations of the previous section were a result of vibrations of the atomic cores. This allowed us to derive many properties, such as the lattice heat capacity and thermal conductivity. Electrons also contribute to this, but in a way that is independent of the lattice, and we can quantify this in the Born–Oppenheimer approximation.

The approximation that we make is that the valence electrons are free to move with relatively few interactions with the ionic cores, and hence we can treat the free electrons as an electron gas. Attempts to do so with classical statistics failed to make accurate predictions. We need to account for the fermionic nature of electrons, in particular for the Pauli exclusion principle.

The assumptions of the free electron Fermi gas model are

- We can treat electrons as a gas of particles subject to the Pauli exclusion principle.
- Between collisions electrons don't interact with other electrons, called the **independent electron approximation**.
- Electrons don't interact with the ions, instead the ions simply restrict the electrons motion by providing an approximately square well, called the **free electron approximation**.
- Collisions are instantaneous.
- Electrons are in thermal equilibrium with their surroundings.

Notice that these are essentially the assumptions of the ideal gas, but with the added requirement of Pauli exclusion principle, and two species, electrons and ions, instead of just one.

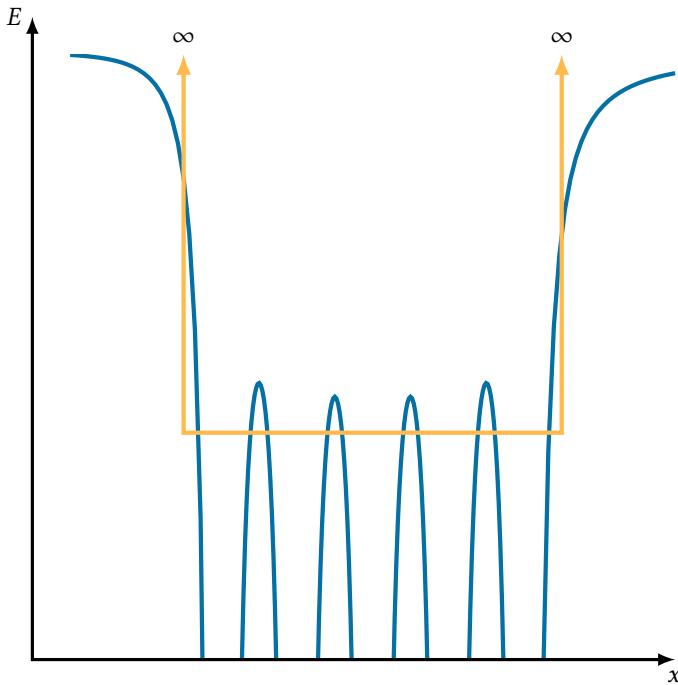


Figure 7.1: The periodic ionic potential in a material can be approximated as an infinite square well potential, where the potential is finite only inside the material.

7.1 Electrons in a Box

7.1.1 One Dimension

In one dimension we can treat the electrons as being in an infinite square well, of width L , produced by the net ionic potential, see Figure 7.1. The independent electron approximation allows us to consider a single electron. The electron's wave function, ψ_n , must satisfy the Schrödinger equation for a free particle inside the box:

$$\hat{\mathcal{H}}\psi_n(x) = -\frac{\hbar^2}{2m} \frac{d^2\psi_n}{dx^2} = \varepsilon_n \psi_n(x). \quad (7.1.1)$$

The boundary conditions being that ψ_n vanishes at 0 and L .

The solutions are

$$\psi_n(x) = A \sin\left(\frac{2\pi}{\lambda_n}x\right), \quad \text{where} \quad \lambda_n = \frac{2L}{n} \quad (7.1.2)$$

with $n \in \mathbb{Z}$, and

$$\varepsilon_n = \frac{\hbar^2}{2m} \left(\frac{n\pi}{L}\right)^2. \quad (7.1.3)$$

Each energy level can be occupied by two electrons, one with spin up and the other with spin down.

7.1.2 Three Dimensions

In three dimensions the same analysis holds, we now consider a three-dimensional infinite square well, which is simply a cube in space which the electrons are constrained to be in. The single electron wave function, $\psi_{\mathbf{k}}$, must satisfy the three-dimensional Schrödinger equation for a free particle inside this box:

$$\mathcal{H}\psi_{\mathbf{k}}(\mathbf{r}) = -\frac{\hbar^2}{2m}\nabla^2\psi_{\mathbf{k}}(\mathbf{r}) = -\frac{\hbar^2}{2m}\left[\frac{d^2}{dx^2} + \frac{d^2}{dy^2} + \frac{d^2}{dz^2}\right]\psi_{\mathbf{k}}(\mathbf{r}) = \varepsilon_{\mathbf{k}}\psi_{\mathbf{k}}(\mathbf{r}). \quad (7.1.4)$$

We could use the standard boundary conditions where $\psi_{\mathbf{k}}$ vanishes on the border of the box, this gives standing wave solutions as in the one-dimensional case. It turns out that it is more useful to use a set of boundary conditions called the **Born-von Karman boundary conditions**. These are a set of periodic boundary conditions in which

$$\psi(\mathbf{r} + L\mathbf{e}_i) = \psi(\mathbf{r}) \quad (7.1.5)$$

where \mathbf{e}_i are unit vectors parallel to the sides of the boxes. We can think of the wave function leaving the box and returning on the opposite side. This gives the box the topology of a 3-torus. This has the advantage of removing the surface of the crystal while keeping the crystal finite.

The solutions are travelling waves, which are just normalised plane waves:

$$\psi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{V}}e^{i\mathbf{k}\cdot\mathbf{r}} \quad (7.1.6)$$

where $V = L^3$ is the volume of the box. The energy eigenvalues are

$$\varepsilon_{\mathbf{k}} = \frac{\hbar^2}{2m}\mathbf{k}^2 = \frac{\hbar^2}{2m}(k_x^2 + k_y^2 + k_z^2). \quad (7.1.7)$$

The wave vector is discretised so that

$$k_i = n_i \frac{2\pi}{L} \quad (7.1.8)$$

for $n_i \in \mathbb{Z}$.

Confusingly we often refer to the states $\psi_{\mathbf{k}}$ as orbitals, although they are not the same as atomic orbitals. These states have discrete energies, $\varepsilon_{\mathbf{k}}$, which depend on the wave vector, \mathbf{k} , which is also discrete. Recall that the momentum operator is defined as $\hat{\mathbf{p}} = -i\hbar\nabla$, and so the momentum is found from the eigenvalue equation

$$\hat{\mathbf{p}}\psi_{\mathbf{k}}(\mathbf{r}) = -i\hbar\nabla\psi_{\mathbf{k}}(\mathbf{r}) = -\frac{i\hbar}{\sqrt{V}}\nabla e^{i\mathbf{k}\cdot\mathbf{r}} = -\frac{i\hbar}{\sqrt{V}}(-i\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}} = \hbar\mathbf{k}\psi_{\mathbf{k}}. \quad (7.1.9)$$

Hence, we find that the momentum is

$$\mathbf{p} = \hbar\mathbf{k}. \quad (7.1.10)$$

We can then define the velocity, by assuming that this momentum is the same as the classical momentum, and so $v = \mathbf{p}/m = \hbar\mathbf{k}/m$. The wavelength of the electron is given by de Broglie's wavelength formula:

$$\lambda = \frac{h}{p} = \frac{2\pi\hbar}{p} = \frac{2\pi}{k} \quad (7.1.11)$$

7.2 Fermi Surface

For a system of N electrons we can populate the allowed energy eigenstates in the order of increasing energy to get the ground state. We can put two electrons in each state, spin up and down, so the first $N/2$ states are occupied. We define the **Fermi energy**, ε_F , as the energy of the highest occupied single electron state in the ground state of the crystal, which is the highest occupied state at 0 K. We define the corresponding wave vector to be the **Fermi wave vector**, k_F . All occupied states then have energy $E \leq \varepsilon_F$ and wave vectors with $|\mathbf{k}| \leq k_F$.

The Fermi wave vector defines a surface in \mathbf{k} -space, which separates occupied and unoccupied states. We call this the **Fermi sphere**, although we will see later that it is not always a sphere.

Recall that in reciprocal space every state has an attached volume $(2\pi/L)^3 = 8\pi^3/V$. This comes from assuming a regular grid of points and working out the volume per point. At $T = 0$ K there is no thermal excitation. Filling the states starting from $\mathbf{k} = 0$ up to $\mathbf{k} = k_F$ we have N electrons. The volume of the Fermi sphere, which is $4\pi k_F^3/3$, must be multiplied by the states per unit volume to give the total number of occupied states, there is also a factor of two since each volume can contain two states. This gives

$$\frac{4}{3}\pi k_F^2 \cdot \left(\frac{8\pi^3}{V}\right)^{-1} \cdot 2 = N. \quad (7.2.1)$$

Rearranging this we have

$$k_F = \left(3\pi^2 \frac{N}{V}\right)^{1/3}. \quad (7.2.2)$$

This is a useful expression as it depends on a single material parameter, $n = N/V$, the electron number density. We can work this out for a single unit cell using the number of free electrons per unit cell and volume of the unit cell.

From this expression we can define several other quantities, first we see that the Fermi energy can be expressed as

$$\varepsilon_F = \frac{\hbar^2}{2m} k_F^2 = \frac{\hbar^2}{2m} \left(3\pi^2 \frac{N}{V}\right)^{2/3}. \quad (7.2.3)$$

We can define the **Fermi temperature**, T_F as

$$\varepsilon_F = k_B T_F, \quad (7.2.4)$$

and the **Fermi velocity** as

$$v_F = \frac{\hbar k_F}{m} = \frac{\hbar}{m} \left(3\pi^2 \frac{N}{V}\right)^{1/3}. \quad (7.2.5)$$

Example 7.2.6 Under standard conditions sodium has a BCC crystal structure with lattice parameter $a = 4.3 \text{ \AA}$ for the conventional unit cell. It is monovalent. Compute the Fermi wave vector, Fermi energy, Fermi temper-

ature, and Fermi velocity. The following may be useful

$$\hbar = 6.58 \text{ eV s}, \quad k_B = 8.62 \times 10^{-5} \text{ eV K}^{-1}, \quad \text{and} \quad m_e = 511 \text{ keV } c^{-2}. \quad (7.2.7)$$

The number of electrons per unit cell is 2, since BCC has two atoms per conventional unit cell and they are monovalent so give one electron each. The volume of the unit cell is $a^3 = (4.3 \text{ \AA})^3$. Combining this

$$n = \frac{N}{V} = \frac{2}{a^3} = \frac{2}{(4.3 \text{ \AA})^3} = \frac{2}{(4.3 \times 10^{-8} \text{ cm})^3} = 2.5 \times 10^{22} \text{ cm}^{-3}. \quad (7.2.8)$$

The Fermi wave vector is then

$$k_F = (3\pi^2 n)^{1/3} = (3\pi^2 \cdot 2.5 \times 10^{22} \text{ cm}^{-3})^{1/3} = 9.1 \times 10^6 \text{ cm}^{-1} = 9.1 \text{ nm}^{-1}. \quad (7.2.9)$$

The Fermi energy is

$$\begin{aligned} \varepsilon_F &= \frac{\hbar^2 k_F^2}{2m_e} = \frac{(6.58 \times 10^{-16} \text{ eV s})^2 (9.1 \times 10^9 \text{ m}^{-1})^2}{2(511 000 \text{ eV } c^{-2})} \\ &= 3.51 \times 10^{-17} \text{ eV } c^2 \text{ s}^2 \text{ m}^{-2} = 3.1 \text{ eV}. \end{aligned} \quad (7.2.10)$$

The Fermi temperature is then

$$T_F = \frac{\varepsilon_F}{k_B} = \frac{3.1 \text{ eV}}{8.62 \times 10^{-5} \text{ eV K}^{-1}} = 36 320 \text{ K}. \quad (7.2.11)$$

The Fermi velocity is

$$\begin{aligned} v_F &= \frac{\hbar k_F}{m_e} = \frac{(6.58 \times 10^{-16} \text{ eV s})(9.1 \times 10^9 \text{ m}^{-1})}{511 000 \text{ eV } c^{-2}} \\ &= 1.17 \times 10^{-11} \text{ s m}^{-1} \text{ } c^{-2} = 1.05 \times 10^6 \text{ m s}^{-1} = 0.0035 \text{ c}. \end{aligned} \quad (7.2.12)$$

Notice how high the Fermi energy is, a classical ideal gas would need to have a temperature of $T_F \approx 36 000 \text{ K}$ to reach these kinetic energies. This is a consequence of Pauli's exclusion principle, which pushes up the energy of electrons just by not allowing them to fall down into already occupied states. All of this occurs even at 0 K, this should make it clear why the classical gas model cannot accurately describe metals.

7.3 Finite Temperatures



For more details on the Fermi–Dirac distribution see the statistical mechanics part of the thermal physics course.

In order to describe the population of states at nonzero temperatures we need to introduce the **Fermi–Dirac distribution**, which is the probability density function describing the distribution occupied states as a function of the energy of the state, ε , the temperature, T , and the **chemical potential**, μ , which is the change in free energy when electrons are added to, or removed from, the system. The Fermi–Dirac distribution is given by

$$f(\varepsilon, \mu, T) = f(\varepsilon) = \frac{1}{e^{(\varepsilon - \mu)\beta} + 1} \quad (7.3.1)$$

where $\beta = 1/(k_B T)$.

From this we can see that $\varphi(\varepsilon = \mu) = 1/2$, which can be taken as the definition of μ . In general μ is temperature dependent. In the limit of $T \rightarrow 0$ we have $\mu \rightarrow \varepsilon_F$, since we recover the 0 K model from before, which is just a step function that is 1 for $\varepsilon < \varepsilon_F$ and 0 for $\varepsilon > \varepsilon_F$. At “normal” temperatures, which is to say $T < T_F$ (not that difficult, T_F tends to be fairly high), we can usually approximate $\mu = \varepsilon_F$ for metals.

! The chemical potential, μ , is sometimes called the **Fermi level**, do not confuse this with the Fermi energy, although both are energies with similar names.

7.3.1 Density of States

Once again we make use of the density of states, $D(\varepsilon) = dN/d\varepsilon$. For a three dimensional free electron Fermi gas if $N(\varepsilon)$ is the total number of states with energy not exceeding ε then $N(\varepsilon_F) = N$ is the total number of electrons. We know

$$\varepsilon_F = \frac{\hbar^2}{2m} k_F^2 = \frac{\hbar^2}{2m} \left(3\pi^2 \frac{N}{V}\right)^{2/3}. \quad (7.3.2)$$

This formula was simply derived by accounting for the volume associated with each state, and so the same applies to any energy, ε , if we replace N with $N(\varepsilon)$.

$$\varepsilon = \frac{\hbar^2}{2m} \left(3\pi^2 \frac{N(\varepsilon)}{V}\right)^{2/3}. \quad (7.3.3)$$

Rearranging this we have

$$N(\varepsilon) = \frac{V}{3\pi^2} \left(\frac{2m}{\hbar^2}\right)^{3/2} \varepsilon^{3/2}. \quad (7.3.4)$$

Therefore

$$D(\varepsilon) = \frac{dN}{d\varepsilon} = \frac{V}{2\pi^2} \left(\frac{2m}{\hbar^2}\right)^{3/2} \sqrt{\varepsilon}. \quad (7.3.5)$$

A useful property therefore, that holds for a free electron gas, not in general, is that

$$D(\varepsilon) = \frac{3}{2} \frac{N(\varepsilon)}{\varepsilon}. \quad (7.3.6)$$

The one further modification that we have to make when using the density of states is that we only want to integrate over the populated states. This means we need to include a factor of $f(\varepsilon, \mu, T)$ in our integrals. For example the number of populated states is

$$N = \int_0^\infty D(\varepsilon) f(\varepsilon, \mu, T) d\varepsilon. \quad (7.3.7)$$

We can use this to determine μ as a function of T by computing the integral and rearranging. We can also compute the total energy stored in the electrons:

$$U_{\text{el}} = \int_0^\infty D(\varepsilon) f(\varepsilon, \mu, T) \varepsilon d\varepsilon. \quad (7.3.8)$$

7.4 Electronic Heat Capacity

We saw in [Section 6.10.2](#) that the phonon contribution to the low temperature heat capacity is proportional to T^3 . In measurements we find that the low temperature heat capacity of metals is modelled by

$$C(T) = \gamma T + AT^3, \quad (7.4.1)$$

where the cubic term is as predicted by Debye's law:

$$C_{\text{phonon}} = \frac{12\pi^4}{5} N k_B \left(\frac{T}{\theta_D} \right)^3. \quad (7.4.2)$$

The linear term must be due to the free electrons in the metal. We will aim to derive the value of γ , which is called the **Sommerfeld constant**, in this section.

7.4.1 Estimate

For a classical gas the heat capacity per particle is

$$C = \frac{3}{2} k_B. \quad (7.4.3)$$

In experiments we measure that the electronic contribution is linear, that is $C_{\text{el}} \propto T$. At room temperature we find that

$$C_{\text{el}} \approx 0.001 \cdot \frac{3}{2} k_B \quad (7.4.4)$$

per electron.

For a free electron gas only electrons close to the Fermi energy, ε_F , can be thermally excited. The fraction of electrons that we consider to be “close” is T/T_F . The kinetic energy of an electron in the free electron gas is

$$U = \frac{3}{2} k_B T. \quad (7.4.5)$$

Therefore if there are N electrons and the fraction that can be excited is T/T_F the kinetic energy of all electrons is

$$U_{\text{el}} \approx \frac{3}{2} N k_B T \frac{T}{T_F} = \frac{3}{2} k_B \frac{T^2}{T_F}. \quad (7.4.6)$$

We then approximate the electronic heat capacity as

$$C_{\text{el}} = \frac{dU_{\text{el}}}{dT} \approx 3Nk_B \frac{T}{T_F}. \quad (7.4.7)$$

We see that C_{el} is lower than the classical value, $3Nk_B T$, by a fraction of $T/T_F \approx 0.01$, so this agrees reasonably well with experimental results.

7.4.2 Electronic Heat Capacity

Consider a free electron gas of N electrons. We assume that we are working at “normal” temperatures, by which we mean $T \ll T_F$, which means $k_B T \ll \varepsilon_F$, and $\mu \approx \varepsilon_F$. The electronic heat capacity is defined as

$$C_{\text{el}} := \frac{dU_{\text{el}}}{dT} \quad (7.4.8)$$

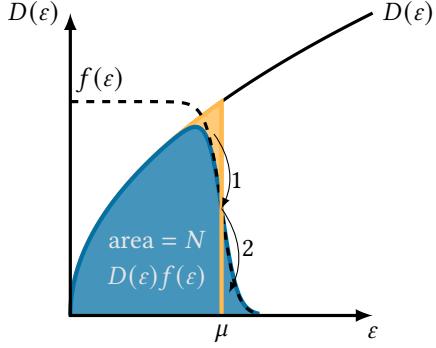


Figure 7.2: In blue is $D(\varepsilon)f(\varepsilon)$ at $T > 0$, in yellow is $D(\varepsilon)f(\varepsilon)$ at $T = 0$, the black line is $D(\varepsilon)$, and the dashed line is $f(\varepsilon)$. When increasing the temperature from 0 to some finite temperature we can do so in two steps, first taking all electrons that will become excited to energy ε_F , and then to their final, higher energy.

where U_{el} is the energy of the electrons. Since the electrons are free in our model the only change in energy can come from changing kinetic energy, which come from thermal excitations.

In order to evaluate the derivative,

$$\frac{dU_{\text{el}}}{dT} \quad (7.4.9)$$

we insert a constant, $U_{\text{el}}(0)$, and define

$$\Delta U := U_{\text{el}}(T) - U_{\text{el}}(0). \quad (7.4.10)$$

We then have

$$\frac{d}{dT}\Delta U = \frac{d}{dT}U_{\text{el}}(T) - \frac{d}{dT}U_{\text{el}}(0) = \frac{d}{dT}U_{\text{el}}(T). \quad (7.4.11)$$

We can consider what happens to the energy by considering two steps. First we excite electrons to energy ε_F , and then to their final energy, note that this is purely theoretical, the exclusion principle forbids exciting all the electrons to the same state at once. See Figure 7.2. We therefore increase the kinetic energy in two steps. For the first step the energy change is

$$\Delta U_1 = \int_0^{\varepsilon_F} [\varepsilon_F - \varepsilon][1 - f(\varepsilon)]D(\varepsilon) d\varepsilon, \quad (7.4.12)$$

where ε is their initial energy and so $\varepsilon_F - \varepsilon$ is the energy needed to excite them to the Fermi energy. We use the distribution $1 - f(\varepsilon)$, since we need to count the states that are no longer occupied after thermal excitation.

For the second step the energy change is

$$\Delta U_2 = \int_{\varepsilon_F}^{\infty} [\varepsilon - \varepsilon_F]f(\varepsilon)D(\varepsilon) d\varepsilon \quad (7.4.13)$$

where ε is the final energy, and so $\varepsilon - \varepsilon_F$ is the increase needed to reach their excited energies from the Fermi energy.

Hence

$$\Delta U = \Delta U_1 + \Delta U_2 \quad (7.4.14)$$

$$= \int_0^{\varepsilon_F} [\varepsilon_F - \varepsilon] [1 - f(\varepsilon)] D(\varepsilon) d\varepsilon + \int_{\varepsilon_F}^{\infty} [\varepsilon - \varepsilon_F] f(\varepsilon) D(\varepsilon) d\varepsilon \quad (7.4.15)$$

$$= \int_0^{\varepsilon_F} [\varepsilon_F - \varepsilon] D(\varepsilon) d\varepsilon + \int_0^{\varepsilon_F} [\varepsilon_F - \varepsilon] f(\varepsilon) D(\varepsilon) d\varepsilon \quad (7.4.16)$$

$$+ \int_{\varepsilon_F}^{\infty} [\varepsilon - \varepsilon_F] f(\varepsilon) D(\varepsilon) d\varepsilon \quad (7.4.17)$$

$$= \int_0^{\infty} [\varepsilon - \varepsilon_F] f(\varepsilon) D(\varepsilon) d\varepsilon + \text{constant}. \quad (7.4.18)$$

The constant here being the first integral, which is constant with respect to temperature, since $f(\varepsilon)$ doesn't appear in the first integral, and this is the only temperature dependent part.

We then have

$$C_{\text{el}} = \frac{dU_{\text{el}}}{dT} \quad (7.4.19)$$

$$= \frac{d}{dT} \Delta U \quad (7.4.20)$$

$$= \frac{d}{dT} \int_0^{\infty} [\varepsilon - \varepsilon_F] f(\varepsilon) D(\varepsilon) d\varepsilon \quad (7.4.21)$$

$$= \int_0^{\infty} [\varepsilon - \varepsilon_F] \frac{df}{dT} D(\varepsilon) d\varepsilon \quad (7.4.22)$$

In order to evaluate this we need to consider f . The Fermi–Dirac distribution is approximately constant at low and high temperatures, and is steeply decreasing in a region of width $2k_B T$ about ε_F . Assuming that $D(\varepsilon)$ varies slowly around $\varepsilon_F \pm k_B T$ we can treat df/dT a bit like a delta distribution and we have

$$C_{\text{el}} = \int_0^{\infty} [\varepsilon - \varepsilon_F] \frac{df}{dT} D(\varepsilon) d\varepsilon \approx D(\varepsilon_F) \int_0^{\infty} [\varepsilon - \varepsilon_F] \frac{df}{dT} d\varepsilon \quad (7.4.23)$$

where the integral now provides normalisation.

This integral is now material independent. In order to calculate the integral we write the Fermi–Dirac distribution as

$$f(\varepsilon, \varepsilon_F, \tau) = \frac{1}{e^{(\varepsilon - \varepsilon_F)/\tau} + 1} \quad (7.4.24)$$

where we will later set $\tau = k_B T$ and we have used $\mu \approx \varepsilon_F$. We therefore have

$$\frac{df}{d\tau} = \frac{\varepsilon - \varepsilon_F}{\tau^2} \frac{e^{(\varepsilon - \varepsilon_F)/\tau}}{[e^{(\varepsilon - \varepsilon_F)/\tau} + 1]^2} = \frac{1}{k_B} \frac{df}{dT}. \quad (7.4.25)$$

Now let $x = (\varepsilon - \varepsilon_F)/\tau = (\varepsilon - \varepsilon_F)/(k_B T)$, and we have

$$C_{\text{el}} = k_B^2 T D(\varepsilon_F) \int_{-\varepsilon_F/\beta}^{\infty} \frac{x^2 e^x}{[e^x + 1]} dx. \quad (7.4.26)$$

For “normal” temperatures the integrand is vanishingly small outside of the region of integration, which is for $x < -\epsilon_F \beta$, we can therefore extend the region of integration down to negative infinity and look up the integral:

$$\int_{-\infty}^{\infty} \frac{x^2 e^x}{[e^x + 1]^2} dx = \frac{\pi^2}{3}. \quad (7.4.27)$$

We then have

$$C_{\text{el}} \approx \frac{1}{3} \pi^2 k_B^2 T D(\epsilon_F). \quad (7.4.28)$$

For a three-dimensional free electron gas recall that

$$D(\epsilon) = \frac{3}{2} \frac{N(\epsilon)}{\epsilon}, \quad (7.4.29)$$

and so

$$D(\epsilon_F) = \frac{3N}{2\epsilon_F} = \frac{3N}{2k_B T_F} \quad (7.4.30)$$

where we have used $N = N(\epsilon_F)$. We therefore have

$$C_{\text{el}} = \frac{1}{2} \pi^2 N k_B \frac{T}{T_F}. \quad (7.4.31)$$

This differs from the estimate in [Equation \(7.4.7\)](#) by only a constant factor, here we have $\pi^2/2 \approx 4.93$ instead of 3. This result is linear in T , and also significantly lower than the classical heat capacity prediction of $C = 3Nk_B/2$.

7.4.3 Thermal Effective Mass

Experimentally if we measure the heat capacity and plot T^2 against C/T we get a straight line described by

$$\frac{C}{T} = \gamma + AT^2. \quad (7.4.32)$$

So, the y -intercept gives the value of γ .

If we do this we find that the value of γ that we have computed,

$$\gamma_{\text{FEG}} = \frac{1}{2} \pi^2 N k_B \frac{1}{T_F}, \quad (7.4.33)$$

using the free electron gas model is not quite equal to the experimental value, γ_{exp} . Of course, this is because the free electron gas is just an approximation, although it does get quite close to the actual value for simple crystals. For example, for potassium, with the BCC crystal structure and lattice parameter $a = 5.23 \text{ \AA}$ we find that $N/V = 2/a^3$ and so

$$T_F = \frac{1}{k_B} \frac{\hbar^2}{2m} \left(3\pi^2 \frac{N}{V} \right)^{2/3} \approx 24\,600 \text{ K}. \quad (7.4.34)$$

We then find that, for one mole ($N = N_A$) of electrons ($m = m_e$), we expect

$$\gamma_{\text{FEG}} = 1.67 \text{ mJ mol}^{-1} \text{ K}^{-2} \quad (7.4.35)$$

which is close to the experimental value of

$$\gamma_{\text{exp}} = 2.08 \text{ mJ mol}^{-1} \text{ K}^{-2}. \quad (7.4.36)$$

For as many approximations as we made in the free electron gas model an error of about 20 % is pretty good.

The Sommerfeld parameter, γ , is proportional to the inverse Fermi temperature, which is in turn proportional to the electron mass. We define a new quantity, called the **thermal effective mass**

$$m_{\text{th}} := m_e \frac{\gamma_{\text{exp}}}{\gamma_{\text{FEG}}}, \quad (7.4.37)$$

which we can think of as being the mass that an electron would have to have for our calculation to match the experimental result.

Typically for simple metals we find that m_{th} is between one and two times the electron mass, but for more complicated metals, such as the d-block Fe, Co, or Ni, the thermal mass can be more like ten or fifteen times the electron mass. For the so called heavy fermion metals, such as CeAl₃, CeCu₂Si₂, and UBe₁₃, with lots of f electrons, we find that the thermal mass can be 100 to 1000 times the electron mass. One explanation for why the electrons appear to be more massive is because we aren't taking into account their interaction with the crystal potential, we will do so later. We also aren't accounting for electron-phonon or electron-electron interactions.

7.5 Electrical Conductivity

Ohm's law states that $I = V/R$ for current, I , voltage, V , and resistance, R . A more general form of Ohm's law in terms of the current density, j , electric field, E , resistivity, ρ , and electrical conductivity, $\sigma = 1/\rho$ (assuming an isotropic medium)

$$j = \frac{1}{\rho} E = \sigma E. \quad (7.5.1)$$

Experimentally we find that for most metals at room temperature or above the resistivity is approximately linear in T .

For a free electron Fermi gas in an electric field we have that the force is given by

$$F = m \frac{dv}{dt} = \hbar \frac{dk}{dt} = -eE. \quad (7.5.2)$$

If there is no scattering then we expect that

$$\delta k(t) = k(t) - k(0) = -\frac{et}{\hbar} E. \quad (7.5.3)$$

From this we would expect that all of the electrons are accelerated with acceleration $\hbar \delta k(t)/m$, this corresponds to the Fermi sphere, and all states within, being shifted by $\delta k(t)$. This is correct for small time scales, but doesn't give a steady-state solution since the velocity becomes arbitrarily large.

The reason for this non-physical behaviour is we haven't considered scattering, which slows the electrons down. We still assume that all electrons have the same

acceleration, but now that they scatter with an average time τ between collisions. Therefore we instead find that

$$\delta\mathbf{k} = -\frac{e\tau}{\hbar}\mathbf{E}. \quad (7.5.4)$$

The average electron velocity, which is called the **drift velocity**, is given by

$$\mathbf{v}_d = \frac{\hbar}{m}\delta\mathbf{k} = -\frac{e\tau}{m}\mathbf{E}. \quad (7.5.5)$$

The current density is then given by the drift velocity times the charge and the charge carrier density,

$$\mathbf{j} = -en\mathbf{v}_d = \frac{ne^2\tau}{m}\mathbf{E} = \sigma\mathbf{E}. \quad (7.5.6)$$

Hence, the electrical conductivity is

$$\sigma = \frac{ne^2\tau}{m}. \quad (7.5.7)$$

This is called the **Drude model**, or sometimes the Drude–Sommerfeld model.

7.5.1 Resistivity Temperature Dependence

In this model the resistivity is

$$\rho = \frac{1}{\sigma} = \frac{m}{ne^2\tau}. \quad (7.5.8)$$

The temperature dependence here enters through the scattering time, τ , which we can relate to the scattering rate, τ^{-1} . At high temperatures, $T > \theta_D$, the scattering rate, τ^{-1} , is proportional to the average phonon density, $\langle n_{ph} \rangle$. This in turn is proportional to T in the high temperature regime, and so $\rho \propto \tau^{-1} \propto \langle n_{ph} \rangle \propto T$. This agrees with experimental observations. We are assuming here that electron-phonon scattering is the dominant scattering type at high temperatures.

We can be more accurate if we consider a variety of types of scattering. There are three types of scattering, electron-phonon scattering, with scattering time τ_{ph} , as discussed above, scattering off of static crystal defects, with scattering time $\tau_{defects}$, and electron-electron scattering, with scattering time τ_{el} . We can define the effective scattering time, τ , to satisfy

$$\frac{1}{\tau} = \frac{1}{\tau_{ph}} + \frac{1}{\tau_{defects}} + \frac{1}{\tau_{el}}. \quad (7.5.9)$$

In the very low temperature regime we see that the resistivity is approximately constant. In this regime both electron-electron and electron-phonon scattering rates go to zero so this must be due to scattering off of static defects.

So, in order to determine the temperature dependence of the scattering rate we need to know the temperature dependence for each type of scattering.

The scattering rate for scattering off of defects is approximately constant, since the number of defects doesn't change with temperature. This is the predominant scattering type at very low temperatures.

At low to intermediate temperatures the temperature dependence is quite complicated. The electron-electron scattering rate depends on the number of electrons

present to scatter, and the number of electrons present to scatter off, and so is proportional to T^2 .

The electron-phonon scattering rate is more complex. Suppose that an electron is scattered from a state with wave vector \mathbf{k} to a state with wave vector \mathbf{k}' , due to an interaction with a phonon with wave vector \mathbf{q} . Conservation of momentum means we must have

$$\mathbf{k}' = \mathbf{k} + \mathbf{q}. \quad (7.5.10)$$

Conservation of energy means that we must have

$$\epsilon' = \epsilon \pm \hbar\omega_{\text{phonon}} \approx \epsilon \quad (7.5.11)$$

where ϵ is the energy of the initial state, ϵ' is the energy of the final state, and $\hbar\omega_{\text{phonon}}$ is the energy of the phonon, which is significantly smaller than the energy of the electrons, since phonon energies are typically a few meV whereas electron energies are a few eV. We say that the scattering is quasi-elastic, since the energy change is mostly negligible.

There are further constraints due to the exclusion principle. First, the electron must scatter from an occupied state, but it must also be close enough to the Fermi energy that it can be scattered, hence we must have $k \lesssim k_F$. The electron must scatter into an unoccupied state, so $k' \gtrsim k_F$. We therefore have

$$k \approx k' \approx k_F. \quad (7.5.12)$$

Essentially meaning that only electrons in states near the Fermi surface can scatter, and that they scatter to states near the Fermi surface.

We will see in a later section that electrons in a periodic potential have periodic wave functions. This means that for every state outside of the first Brillouin zone there is an equivalent state in the first Brillouin zone. For this reason we get normal and umklapp scattering, just as with phonons.

In normal scattering the change in direction of travel is small. For multiple normal scattering events the net effect tends to cancel and therefore normal scattering has a negligible contribution to electrical resistivity, apart from at very low temperatures when there aren't many phonons with enough energy to scatter an electron into the next Brillouin zone.

In umklapp scattering we can reduce $\mathbf{k} + \mathbf{q}$ to a wave vector in the first Brillouin zone by addition of a lattice vector, \mathbf{G} , such that $\mathbf{k}' = \mathbf{k} + \mathbf{q} + \mathbf{G}$ is in the first Brillouin zone. Typically these electrons will end up in states on the opposite side of the Brillouin zone, having only just entered the neighbouring Brillouin zone. For this reason the direction change in these types of scattering events is large. Therefore there is a significant contribution to the electrical resistivity from umklapp scattering, apart from at very low temperatures when there aren't enough energetic phonons.

In order for umklapp scattering to happen we need phonons with $q > q_0$, where q_0 is the difference in wave vectors between the closest points on Fermi surfaces in neighbouring Brillouin zones. The number of phonons that satisfy this is given by a Boltzmann factor, $e^{-\vartheta_U/T}$, where ϑ_U is a characteristic material property, with dimensions of temperature. Therefore the electron-phonon scattering at low to intermediate temperatures is proportional to $e^{-\vartheta_U/T}$.

To summarise the temperature dependence of the scattering rate:

- At very low temperatures, $T \ll \vartheta_D$, resistivity approaches a constant value as most scattering occurs due to defects in the crystal.

- At low to intermediate temperatures the temperature dependence is more complex. In particular the contribution from electron-electron scattering scales as T^2 and the contribution from electron-phonon scattering scales as $e^{-\vartheta_U/T}$.
- At high temperatures, $T > \vartheta_D$, the scattering rate is proportional to T , since electron-phonon scattering is dominant and there are plenty of phonons.

7.6 Electronic Thermal Conductivity

In a metal electrons have a non-negligible contribution to the transport of heat. The overall thermal conductivity, κ , is given by $\kappa = \kappa_{el} + \kappa_{ph}$, where κ_{el} and κ_{ph} are the contributions due to electrons and phonons respectively.

For metals the thermal conductivity is lowest at low temperatures, it then increases with temperature up to a point, then starts decreasing again as umklapp scattering becomes more important.

We will use the same model for electronic thermal conductivity as we did for phononic thermal conductivity in [Section 6.13](#), namely that

$$\kappa_{el} = \frac{1}{3} C_{el} v \ell \quad (7.6.1)$$

where v is the particle velocity, ℓ is the mean free path, and C_{el} is the electronic heat capacity, which is given by

$$C_{el} = \frac{1}{V} \frac{1}{2} \pi^2 N k_B \frac{T}{T_F}. \quad (7.6.2)$$

The Fermi temperature is given by

$$T_F = \frac{\epsilon_F}{k_B} = \frac{\frac{1}{2} m v_F^2}{k_B}. \quad (7.6.3)$$

The mean free path is given by

$$\ell \approx v_F \tau. \quad (7.6.4)$$

We also define the electron density to be $n = N/V$.

Putting this all together we find that

$$\kappa_{el} = \frac{\pi^2 k_B^2 n}{3m} T \tau. \quad (7.6.5)$$

At high temperatures this is approximately linear in T . At lower temperatures the temperature dependence becomes more complicated as we have to consider the temperature dependence of τ .

It is possible to relate the thermal conductivity to the electrical conductivity, which we calculated earlier was

$$\sigma = \frac{ne^2 \tau}{m}. \quad (7.6.6)$$

To do so we compare their ratio:

$$\frac{\kappa_{el}}{\sigma} = \frac{\pi^2}{3} \left(\frac{k_B}{e} \right)^2 T = L_{FEG} T. \quad (7.6.7)$$

Here we define the **Lorenz number**, L , to be

$$L = \frac{\kappa_{\text{el}}}{\sigma T}. \quad (7.6.8)$$

This is approximately constant, a fact referred to as the **Widemann–Franz law**. For the free electron gas model we have

$$L_{\text{FEG}} = 2.44 \times 10^{-8} \text{ W } \Omega \text{ K}^{-2}. \quad (7.6.9)$$

This turns out to be quite close to experimental values, which is quite impressive for a model with so many approximations.

7.7 Magnetic Field Effects

For free electrons in an electric field, E , and magnetic field, B , we expect the force due to these fields to be given by

$$F = m \frac{dv}{dt} = \hbar \frac{dk}{dt} = -e(E + v \times B). \quad (7.7.1)$$

However, we don't have free electrons, in particular we have to account for scattering. We do so by considering another term $\hbar \delta k / \tau$, for the same reason that we did when considering the electrical conductivity. The force is instead

$$F_d = \hbar \left(\frac{d}{dt} + \frac{1}{\tau} \right) \delta k = -e(E + v_d \times B). \quad (7.7.2)$$

Here F_d is the average force on an electron and $v_d = \hbar \delta k / m$ is the drift velocity, which is just the average velocity.

Using $mv_d = \hbar \delta k$ we have

$$m \left(\frac{d}{dt} + \frac{1}{\tau} \right) v_d = -e(E + v_d \times B). \quad (7.7.3)$$

We will consider only a steady-state solution, that is a solution in which the time derivative vanishes. The result therefore is that the entire Fermi sphere is displaced by δk

7.7.1 Hall Effect

Consider a cuboidal metallic bar in an electromagnetic field, aligned such that the electric field is along the bar, which we will call the x -direction, and the magnetic field acts across the bar, which we will call the z -direction. The electric field, E_x , generates a current density, j_x , along the bar. This corresponds to charges moving along the x -direction. The Lorentz force then has a component, due to the $v \times B$ term, along the y -direction, since at the instant when the magnetic field is switched on, $v = v_x \hat{x}$ and $B = B_z \hat{B}$. The result is that the electrons also move in the y -direction, and hence there is a slight build up of negative charge along one side and a resulting positive charge on the other side. We can measure this charge separation as a voltage, called the **Hall voltage**. The process producing this transverse voltage is called the **Hall effect**. We can now work out what we expect the voltage to be for a given metal.

Considering collisions in the metal the equation of motion is

$$m \left(\frac{d}{dt} + \frac{1}{\tau} \right) \mathbf{v} = -e(\mathbf{E} + \mathbf{v} \times \mathbf{B}), \quad (7.7.4)$$

where we drop the subscript d, but are still using the drift velocity. The magnetic field is $\mathbf{B} = (0, 0, B)$ and so we find that

$$m \left(\frac{d}{dt} + \frac{1}{\tau} \right) v_x = -e(E_x + Bv_y), \quad (7.7.5)$$

$$m \left(\frac{d}{dt} + \frac{1}{\tau} \right) v_y = -e(E_y - Bv_x), \quad (7.7.6)$$

$$m \left(\frac{d}{dt} + \frac{1}{\tau} \right) v_z = -eE_z. \quad (7.7.7)$$

$$(7.7.8)$$

For a steady-state solution the time derivative vanishes and we are left with

$$v_x = -\frac{e\tau}{m} E_x - \omega_c \tau v_y, \quad (7.7.9)$$

$$v_y = -\frac{e\tau}{m} E_y + \omega_c \tau v_x, \quad (7.7.10)$$

$$v_z = -\frac{e\tau}{m} E_z. \quad (7.7.11)$$

Here ω_c is the **cyclotron frequency**, which is the frequency associated with the

¹see the notes on the electromagnetism course for details.

$$\omega_c = \frac{eB}{m}. \quad (7.7.12)$$

Further, in a steady-state since electrons don't leave the bar we must have that $v_y = 0$. Putting this in and combining the first two equations for v_i we have

$$E_y = \frac{m\omega_c\tau}{e\tau} v_x = \frac{m\omega_c}{e} \left(-\frac{e\tau}{m} E_x \right) = -\omega_c \tau E_x = -\frac{eB\tau}{m} E_x. \quad (7.7.13)$$

We can now work out the Hall voltage, V_H , which is the transverse voltage measured along the y -direction. Over a distance d the voltage is given by $V = Ed$. Using this the Hall voltage is

$$V_H = E_y w_y, \quad (7.7.14)$$

where w_y is the width of the bar in the y -direction. Let w_z be the width of the bar in the z -direction, then, using the fact that the current, I , and current density, j , are related by $j = I/A$ we can calculate the longitudinal current:

$$I_x = w_y w_z j_x. \quad (7.7.15)$$

We define the **Hall coefficient**, R_H , to be

$$R_H = \frac{V_H/w_y}{I_x B / (w_y w_z)} = \frac{V_H w_z}{I_x B}. \quad (7.7.16)$$

We can measure all of the quantities on the right fairly easily. We also know that

$$j_x = \sigma E_x = \frac{ne^2\tau}{m} E_x, \quad (7.7.17)$$

for a free electron gas. Substituting this and [Equation \(7.7.13\)](#) into the definition of the Hall coefficient we have

$$R_H = \frac{E_y}{j_x B} = -\frac{eB\tau}{m} E_x \frac{m}{ne^2\tau E_x B} = -\frac{1}{ne}. \quad (7.7.18)$$

So, since e is known, we can use the Hall coefficient as a measure of the electron density of the metal, and we can also measure the Hall coefficient just by measuring voltages, lengths, currents, and magnetic field strengths, all of which can be done fairly easily.

Eight

Electron Energy Bands

The free electron gas model works well for most metals. In order deal with the electronic properties of insulators and semiconductors we need to introduce the concept of electron energy bands, which explain some of the deviation from the free electron gas model. These are energy-wave vector dispersion relations for the electrons. We no longer treat electrons as a free electron gas, but we still assume the independent electron approximation holds.

8.1 Energy Bands

Consider a diatomic molecule. The bonding of such a molecule can be represented by the molecular orbital diagrams such as in [Figure 5.5](#). This gives a minimum energy for the electrons, if they are in the bonding orbital, and a maximum energy, if they are in the antibonding orbital.

If we now consider many atoms of the same type then electrons involved in bonding will have more energy bands, over roughly the same range of energies. In the limit of a very large number of atoms the allowed energy levels become approximately continuous between the minimum and maximum values, and we call the region of allowed energies an **energy band**.

8.2 Band Gaps

In a real metal the valence electrons aren't actually free, but move in a periodic crystal potential. This is given by the superposition of individual Coulomb potentials for the ionic cores, see [Figure 7.1](#). The potential is typically quite weak, essentially the effect is to modify the parabolic free-electron energy dispersion, $\epsilon(\mathbf{k}) = \hbar^2 k^2 / (2m)$, and create band gaps, which are regions with no energy levels. This occurs at the Brillouin zone edges. The electronic properties of solids then depend on where the highest filled energy level is relative to the positions of band gaps.

[Figure 8.1](#) shows approximately how bands are filled for different types of materials. Metals have partially filled bands, which means that only a very small amount of energy is required to excite a valence electron into a slightly higher energy state, where it can be accelerated by an electric field, which results in a current. The size of the band gap is significantly larger than the gain in kinetic energy of a single electron due to an applied voltage, so electrons cannot be excited into the next band simply by applying a voltage.

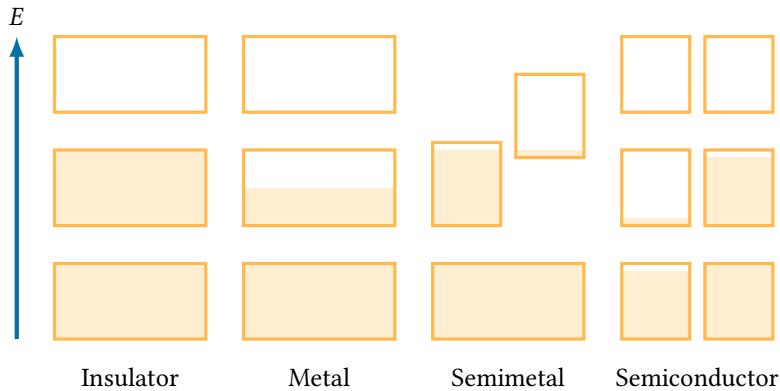


Figure 8.1: The relation between electrical behaviour relates and how full the bands are relative to the positions of band gaps.

Insulators are actually very similar in band structure to metals, except that the bands are filled, and so the electrons cannot be excited by a voltage as they would have to move into the next band, and there isn't enough energy for this.

Semiconductors are somewhere between, at absolute zero they behave as insulators. As temperature increases more and more electrons are excited thermally into the next band, which means that they become better conductors as temperature increases.

Semimetals have bands that overlap, we will discuss this, and semiconductors, in the future.

8.3 Nearly Free Electron Model

To understand band splitting we start from the free electron model with plane wave states, which has no band splitting, and we introduce a periodic potential. The free electron wave functions are

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k} \cdot \mathbf{r}}. \quad (8.3.1)$$

The energy of such a state is

$$\varepsilon_{\mathbf{k}} = \frac{\hbar^2}{2m} k^2 = \frac{\hbar^2}{2m} (k_x^2 + k_y^2 + k_z^2). \quad (8.3.2)$$

The allowed values of k_i are $2\pi n/L$ for $n \in \mathbb{N}$ and L being the length of the crystal. The momentum of such a state is $\mathbf{p} = \hbar\mathbf{k}$.

Recall that the diffraction condition $\Delta\mathbf{k} = \mathbf{G}$, for some reciprocal lattice vector, \mathbf{G} , or equivalently $2\mathbf{k} \cdot \mathbf{G} = \mathbf{G}^2$, arises when we consider plane wave electrons in exactly the same way that it arises when considering x-ray diffraction. Or, in one dimension, $k = \pm G/2 = \pm n\pi/a$. A right-travelling electron with $k = \pi/a$ is diffracted into a left-travelling electron with $k = -\pi/a$, and vice versa. Notice that these are the same state since k^2 is unchanged and both direction and wave vector are negated so $\mathbf{k} \cdot \mathbf{r}$ is unchanged.

Still in one dimension we can write our wave function as

$$\psi(\pm) = \psi_{+k}(x) \pm \psi_{-k}(x), \quad (8.3.3)$$

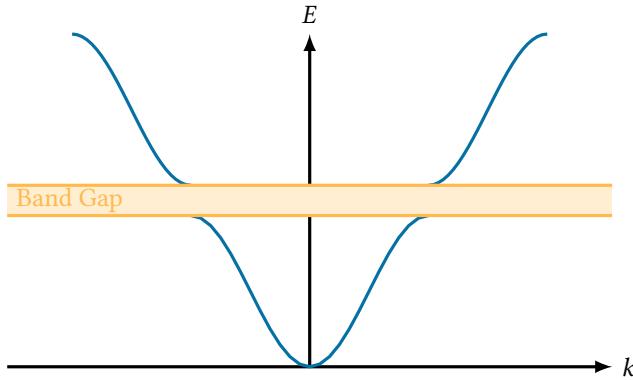


Figure 8.2: Energy bands separated by a band gap.

which gives standing wave solutions

$$\psi(+)=2\cos\left(\frac{\pi x}{a}\right), \quad \text{and} \quad \psi(-)=2i\sin\left(\frac{\pi x}{a}\right). \quad (8.3.4)$$

The probability density is defined as $\rho = \psi^*\psi = |\psi|^2$. For travelling waves we have

$$\rho = e^{ikx}e^{-ikx} = 1. \quad (8.3.5)$$

For the standing waves we have

$$\rho(+) = |\psi(+)|^2 \propto \cos^2\left(\frac{\pi x}{a}\right), \quad \text{and} \quad \rho(-) = |\psi(-)|^2 \propto \sin^2\left(\frac{\pi x}{a}\right). \quad (8.3.6)$$

These have maxima at $x = na$ and $x = (2n + 1)a/2$ respectively for $n \in \mathbb{Z}$. This is pretty important since the ionic cores are placed at the positions na so the maxima of the + standing waves align with the ions, and the maxima of the - standing waves are directly between. The result is, since the potential energy scales as $-1/r^2$, the + states are lower energy than the - states. We can therefore associate the + states with the lower band and the - states with the higher band. The bands, and the forbidden region between them, are shown in [Figure 8.2](#).

So we know now why the splitting occurs, but not how large the splitting is. For this we will need to consider the potential. To aid our thinking we consider a simple potential with the same periodicity as the crystal potential, namely

$$U(x) = -U \cos\left(\frac{2\pi x}{a}\right). \quad (8.3.7)$$

At the zone boundary we have the, now normalised, wave functions

$$\psi(+) = \sqrt{\frac{2}{a}} \cos\left(\frac{\pi x}{a}\right), \quad \text{and} \quad \psi(-) = \sqrt{\frac{2}{a}} \sin\left(\frac{\pi x}{a}\right). \quad (8.3.8)$$

The band splitting, E_g , is simply the difference between the average energies of the

two states, which is

$$E_g = \varepsilon_- - \varepsilon_+ \quad (8.3.9)$$

$$= \int_0^a U(x) |\psi(-)|^2 dx - \int_0^a U(x) |\psi(+)|^2 dx \quad (8.3.10)$$

$$= \int_0^a U(x) (|\psi(-)|^2 - |\psi(+)|^2) dx \quad (8.3.11)$$

$$= -U \frac{2}{a} \int_0^a \cos\left(\frac{2\pi x}{a}\right) \left[\sin^2\left(\frac{\pi x}{a}\right) - \cos^2\left(\frac{\pi x}{a}\right) \right] dx \quad (8.3.12)$$

$$= U \frac{2}{a} \int_0^a \cos^2\left(\frac{2\pi x}{a}\right) dx \quad (8.3.13)$$

$$= \frac{U}{a} \int_{-a}^a \cos^2\left(\frac{2\pi x}{a}\right) dx \quad (8.3.14)$$

$$= U \quad (8.3.15)$$

Here we have used the trig identity

$$\sin^2 \vartheta - \cos^2 \vartheta = -\cos(2\vartheta). \quad (8.3.16)$$

We then noticed the similarity to computing Fourier components, in particular that this integral gives a , since \cos is orthogonal in $L^2([-a, a])$, and has norm a .

While this simple potential isn't realistic the real potential is periodic and we can always choose to have an ion at 0, and so the potential will be an even function. We can therefore expand it as a Fourier series. Doing so, and generalising to three dimensions, we find that for a realistic crystal potential, U , the splitting at wave vector \mathbf{k} in the Brillouin zone is given by the Fourier component, $U_{2\mathbf{k}}$, of the crystal potential. Recall that the Fourier components are only non-zero for wave vectors which coincide with reciprocal lattice vectors, hence band splittings can appear only in locations where $2\mathbf{k} = \mathbf{G}$ for some reciprocal lattice vector \mathbf{G} . In particular this requirement is met at the centre of the faces of a three-dimensional Brillouin zone, since by construction the Brillouin zone faces are halfway between reciprocal lattice points, which are separated by a reciprocal lattice vector.

While the actual electronic band structure for a realistic crystal potential is too complicated to compute analytically there are several model potentials that exhibit important behaviour. We've already seen the cosine model, another model potential is the **Kronig–Penny model**, which is simply a square wave.

8.4 Empty Lattice Approximation

While the ionic potential is important for a lot of materials the main splitting is due to the lattice structure. We therefore consider the **empty lattice approximation**, in which we take $U = 0$ and consider how the lattice periodicity effects the quadratic free electron dispersion relation,

$$\varepsilon_{\mathbf{k}} = \frac{\hbar^2}{2m} k^2. \quad (8.4.1)$$

The lattice is periodic, in particular $\varepsilon(\mathbf{k}) = \varepsilon(\mathbf{G} + \mathbf{k})$ for any reciprocal lattice vector \mathbf{G} . There are two approaches we can take to account for this. One is to consider

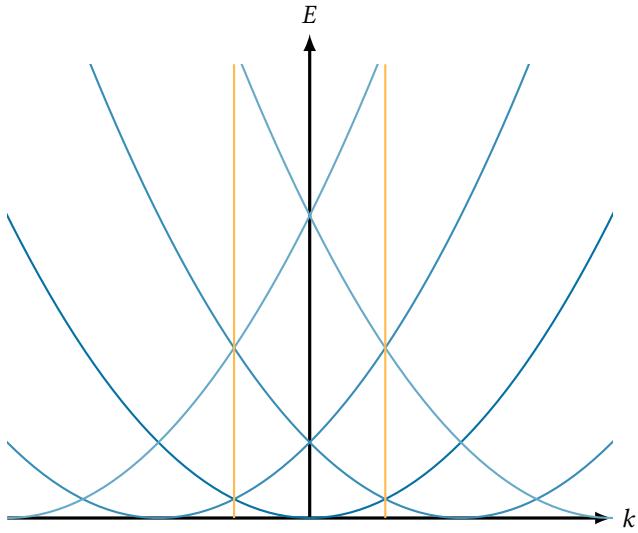


Figure 8.3: ε_k and all shifts of ε_k by some lattice vector G .

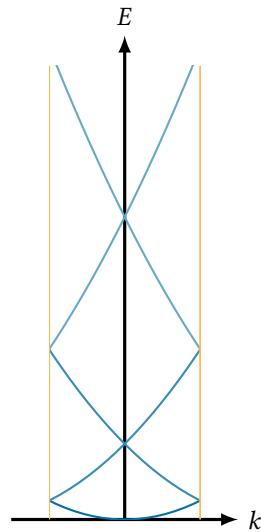


Figure 8.4: ε_k and all G -shifts of ε_k for some reciprocal lattice vector G .

ε_k , and all shifts of the dispersion relation by a reciprocal lattice vector, discarding everything outside of the first Brillouin zone. The second is to consider ε_k and map all parts outside of the Brillouin zone into the Brillouin zone by addition of an appropriate lattice vector. We consider both approaches in one dimension but the first is far simpler for higher dimensions.

In one dimension the dispersion relation, and all G -shifts of ε_k , are shown in Figure 8.3. Restricting the result to just the Brillouin zone we get Figure 8.4.

The alternative method, of shifting the parts of ε_k outside of the Brillouin zone by some reciprocal lattice vector such that they are inside the Brillouin zone, is illustrated in Figure 8.5.

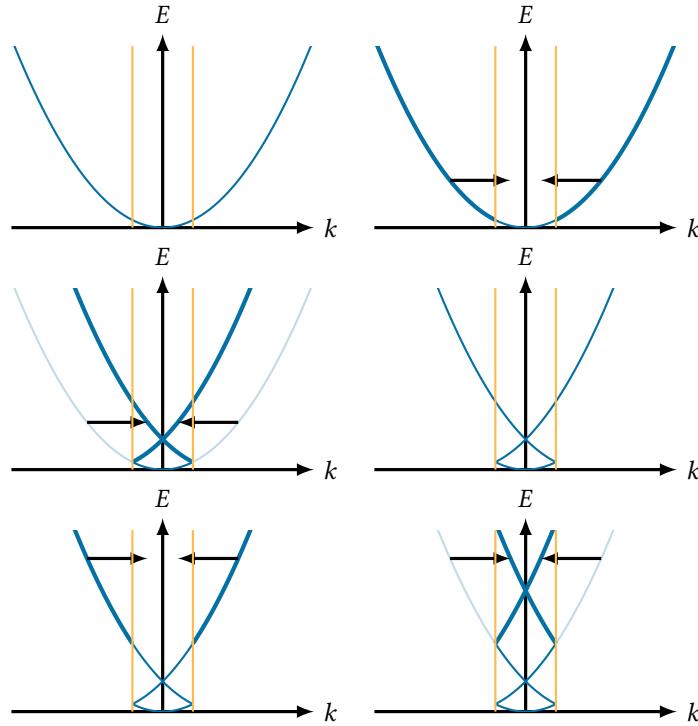


Figure 8.5: The process for mapping a dispersion relation outside of the Brillouin zone into the Brillouin zone by mapping parts outside by a reciprocal lattice vector.

In more dimensions we can do exactly the same, although the first method of shifting is much easier to visualise. The difference is that instead of parabolas we use paraboloids. We then typically plot the intersection of these paraboloids with some plane. For example plotting for the (y, z) -plane we get the result shown in Figure 8.6. To see where the second, higher up parabola comes in see Figure 8.7 which shows how a parabola shifted by a lattice vector perpendicular to the plane intersects the plane to give a higher up parabola.

Consider the equation for the shifted paraboloids in three dimensions. We have

$$\varepsilon(\mathbf{k}) = \frac{\hbar^2}{2m}(\mathbf{k} + \mathbf{G})^2 = \frac{\hbar^2}{2m}[(k_x + G_x)^2 + (k_y + G_y)^2 + (k_z + G_y)^2]. \quad (8.4.2)$$

We can now go systematically through the lattice vectors and work out all of the paraboloids that we need to consider. First we have $\mathbf{G} = \mathbf{0}$, which gives us the base paraboloid

$$\varepsilon(\mathbf{k}) = \frac{\hbar^2}{2m}k^2 = \frac{\hbar^2}{2m}[k_x^2 + k_y^2 + k_z^2]. \quad (8.4.3)$$

In the (k_y, k_z) -plane we have $k_y = k_z = 0$ and so this gives the parabola

$$\varepsilon(k_x) = \frac{\hbar^2}{2m}k_x^2. \quad (8.4.4)$$

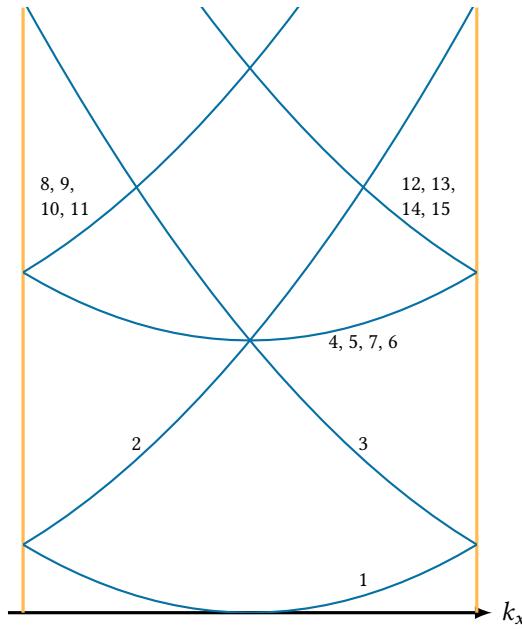


Figure 8.6: Electron bands in two dimensions.

Then we can consider $G = \mathbf{a}_1^*$, which points in the “ x ”-direction in reciprocal space. For a simple cubic lattice we then have

$$\varepsilon(\mathbf{k}) = \frac{\hbar^2}{2m}(\mathbf{k} + \mathbf{a}_1^*)^2 = \frac{\hbar^2}{2m}[(k_x + 2\pi/a)^2 + k_y^2 + k_z^2]. \quad (8.4.5)$$

In the (k_y, k_z) -plane we then have

$$\varepsilon(\mathbf{k}) = \frac{\hbar^2}{2m}(k_x + 2\pi/a)^2. \quad (8.4.6)$$

Next we consider $G = \mathbf{a}_2^*$, we then have

$$\varepsilon(\mathbf{k}) = \frac{\hbar^2}{2m}(\mathbf{k} + \mathbf{a}_2^*)^2 = \frac{\hbar^2}{2m}[k_x^2 + (k_y + 2\pi/a)^2 + k_z^2]. \quad (8.4.7)$$

In the (k_y, k_z) -plane we then have

$$\varepsilon(k_x) = \frac{\hbar^2}{2m}[k_x^2 + (2\pi/a)^2]. \quad (8.4.8)$$

The case for $G = \mathbf{a}_3^*$ is the same. Repeating this for all $G = v_i \mathbf{a}_i^*$ for $v_i \in \mathbb{Z}$ we can build up the results in Table 8.1.

8.5 Bloch's Theorem

The periodicity of a crystal lattice imposes constraints on possible electron wave functions. In particular **Bloch's theorem** states that all solutions to Schrödinger's equation must be of the form

$$\psi_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}} \quad (8.5.1)$$

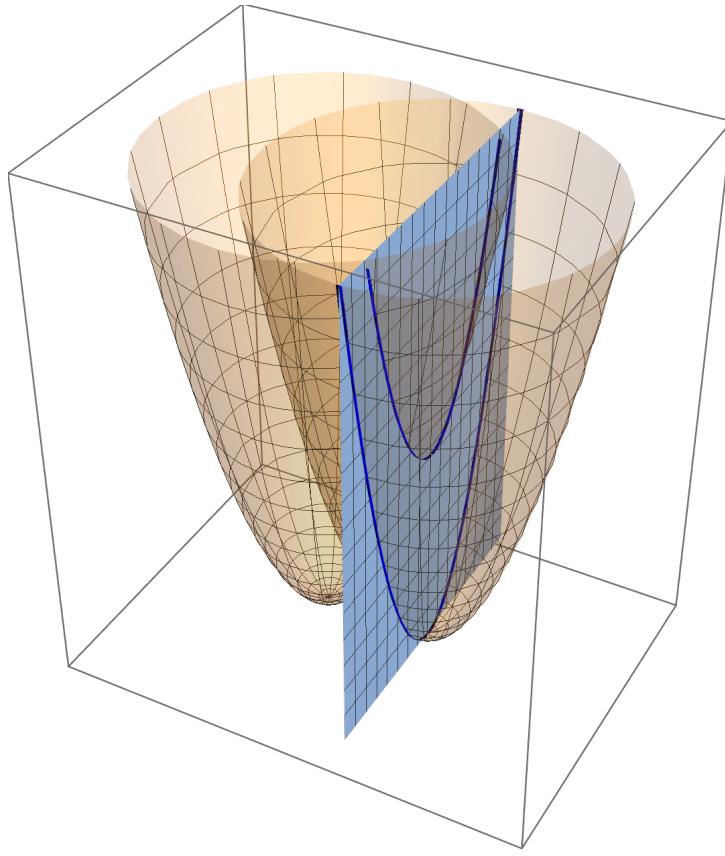


Figure 8.7: A paraboloid at the origin and a shifted paraboloid intersect a plane along two parabolas.

Table 8.1: The value of ε at $\mathbf{k} = \mathbf{0}$ and $\mathbf{k} = k_x \mathbf{a}_1^*$ for all G -shifts with $|G| \leq \sqrt{2}(2\pi/a)$. Taking $\hbar^2/(2m) = 1$ for simplicity.

Band	G	$\varepsilon(000)$	$\varepsilon(k_x 00)$
1	000	0	k_x^2
2	100	$(2\pi/a)^2$	$(k_x + 2\pi/a)^2$
3	$\bar{1}00$	$(2\pi/a)^2$	$(k_x - 2\pi/a)^2$
4–6	010, 0 $\bar{1}$ 0, 001, 00 $\bar{1}$	$(2\pi/a)^2$	$k_x^2 + (2\pi/a)^2$
8–11	110, 101, 1 $\bar{1}$ 0, 10 $\bar{1}$	$2(2\pi/a)^2$	$(k_x + 2\pi/a)^2 + (2\pi/a)^2$
12–15	1 $\bar{1}$ 0, 101, 1 $\bar{1}$ 0, 10 $\bar{1}$	$2(2\pi/a)^2$	$(k_x - 2\pi/a)^2 + (2\pi/a)^2$
16–19	011, 01 $\bar{1}$, 0 $\bar{1}$ 1, 0 $\bar{1}$ $\bar{1}$	$2(2\pi/a)^2$	$k_x^2 + 2(2\pi/a)^2$

where $u_{\mathbf{k}}$ is some function with the same periodicity as the lattice. This means that $u_{\mathbf{k}}(\mathbf{r} + \mathbf{T}) = u_{\mathbf{k}}(\mathbf{r})$, where \mathbf{T} is a translation vector of the lattice.

This result means that the possible one electron wave functions, $\psi_{\mathbf{k}}$, are plane waves, $e^{i\mathbf{k}\cdot\mathbf{r}}$, modulated by periodic functions, $u_{\mathbf{k}}$. This does *not* mean that the wave functions are periodic.

The consequence of this theorem is that $\psi_{\mathbf{k}+\mathbf{G}}(\mathbf{r}) = \psi_{\mathbf{k}}(\mathbf{r})$ and $\varepsilon(\mathbf{k} + \mathbf{G}) = \varepsilon(\mathbf{k})$, that is the wave function and energy are periodic in the wave vector for any reciprocal lattice vector, \mathbf{G} .

Bloch's theorem means that we only need to consider electron states where the wave vectors lie inside the first Brillouin zone. For a state with a wave vector, \mathbf{k}' , outside the first Brillouin zone there is an equivalent state with wave vector $\mathbf{k} = \mathbf{k}' - \mathbf{G}$ for some suitable reciprocal lattice vector, \mathbf{G} , such that \mathbf{k} is in the first Brillouin zone.

8.5.1 Bloch's Theorem in One Dimension

Consider a system of N lattice points arranged in a ring. Suppose there is a periodic potential, U , such that $U(x + a) = U(x)$ where x is the distance around the ring and a is the distance between lattice points around the ring. The probability density function for an electron must be periodic around the ring, meaning $|\psi(x + a)|^2 = |\psi(x)|^2$, which is required as the system is invariant under a rotation taking each lattice point to its neighbour. This implies that $\psi(x + a) = C\psi(x)$ where $C \in U(1)$ is a complex number with unit modulus. From this we see that ψ is almost periodic, it just changes phase.

If we go once around the ring we get back to where we started. The only physically reasonable thing to expect is that the wave function takes the same value after going once around the ring, so $\psi(x + Na) = \psi(x) = C^N\psi(x)$. From this we see that $C = e^{2\pi s/N}$, where $s = 0, 1, \dots, N - 1$. That is C is the N th root of unity.

The most general solution satisfying these constraints is that

$$\psi(x) = u_k(x)e^{i2\pi(x/a)(s/N)} \quad (8.5.2)$$

where the extra factor of x/a allows us to consider positions between lattice points. Here u_k is some periodic function such that $u_k(x) = u_k(x + a)$. We identify $k = 2\pi s/(Na) = 2\pi/L$ where L is the circumference of the ring. We see that we get periodic wave vectors.

8.5.1.1 A Comment on the Wave Vector

The wave vector, \mathbf{k} , labels a Bloch wave function, that is \mathbf{k} is a quantum number. In the free electron limit $\hbar\mathbf{k}$ is the linear momentum. In the non-free electron case (i.e. reality) $\hbar\mathbf{k}$ is *not* the linear momentum. Instead we call $\hbar\mathbf{k}$ the crystal momentum. The wave vector plays the same role as for phonons, for example, it appears in the scattering of an electron by some phonon:

$$\mathbf{k}_{\text{incident}} + \mathbf{q}_{\text{phonon}} = \mathbf{k}_{\text{final}} + \mathbf{G}. \quad (8.5.3)$$

8.6 Schrödinger Equation for a Periodic Potential

Consider a periodic crystal potential, U , we continue to make the independent electron approximation and as such U accounts for the ions and the other electrons and

we only need to solve the Schrödinger equation for a single electron. For simplicity we work in one dimension, but higher dimension cases aren't that different.

Consider some general wave vector, \mathbf{k} , anywhere in the first Brillouin zone. The crystal potential is periodic and as such we can expand it as a Fourier series:

$$U(x) = \sum_G U_G e^{iGx}. \quad (8.6.1)$$

Since the potential is a real function it can be shown that $U_{-G} = U_G^*$.

The Schrödinger equation, $\mathcal{H}\psi = \varepsilon\psi$, then takes the form

$$\left(\frac{\hat{p}^2}{2m} + U(x) \right) \psi(x) = \left(\frac{\hat{p}^2}{2m} + \sum_G U_G e^{iGx} \right) \psi(x) = \varepsilon\psi(x). \quad (8.6.2)$$

We now express the electron wave function as a Fourier *transform*, not series, accounting for the fact that the transform variable, k , is discrete, since $k = 2\pi s/L$ for $s \in \mathbb{Z}$, where $L = Na$ is the length of the crystal. This means we write ψ as

$$\psi(x) = \sum_k C_k e^{ikx}. \quad (8.6.3)$$

Since we are summing over the relatively dense set of wave vectors, rather than over reciprocal lattice vectors, G , this does not yield a periodic wave function. We also include wave vectors outside of the first Brillouin zone.

Considering the first term of the Schrödinger equation we now have

$$\frac{\hat{p}^2}{2m} \psi(x) = \frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2} = -\frac{\hbar^2}{2m} \sum_k k^2 C_k e^{ikx}. \quad (8.6.4)$$

The second term gives

$$\sum_G \sum_k U_G C_k e^{i(G+k)x}. \quad (8.6.5)$$

Renaming k to k' in this term and putting the terms together the Schrödinger equation becomes

$$\sum_k \frac{\hbar^2}{2m} k^2 C_k e^{ikx} + \sum_G \sum_k U_G C_{k'} e^{i(k'+G)x} = \sum_k \varepsilon C_k e^{ikx}. \quad (8.6.6)$$

We can rewrite second term to be in terms of $k = k' + G$ and we get

$$\sum_k \frac{\hbar^2}{2m} k^2 C_k e^{ikx} + \sum_G \sum_k U_G C_{k-G} e^{ikx} = \sum_k \varepsilon C_k e^{ikx}. \quad (8.6.7)$$

Each Fourier component, e^{ikx} , must have the same coefficient on either side of the equation and so we conclude that

$$\frac{\hbar^2}{2m} k^2 C_k + \sum_G U_G C_{k-G} = \varepsilon C_k \quad (8.6.8)$$

for all k .

Defining $\lambda_k := \hbar^2 k^2 / (2m)$ lets us write this as

$$(\lambda_k - \varepsilon)C_k + \sum_G U_G C_{k-G} = 0. \quad (8.6.9)$$

We call this the **central equation**.

We see that we end up with a system of simultaneous linear equations for C_k . We can write this with matrices and the problem of finding the values of C_k becomes a matrix inversion problem. This can then be done efficiently on a computer. To do so we obviously need to truncate the sums at some finite number of terms, keeping enough terms that we achieve the desired accuracy, but not so many that the computation takes too long.

We can reasonably compute the sums keeping enough terms that the limiting factor in accuracy is not the number of terms but the independent electron approximation. Overcoming this barrier is an area of active research. Currently techniques include accounting for some electron interaction in the potential, this is done in the frame work of density functional theory, but this is beyond the scope of this course.

8.6.1 Bloch's Theorem Again

We can “prove” Bloch’s theorem from the central equation. We do not, in general, expect that the wave function, ψ , given by

$$\psi(x) = \sum_k C_k e^{ikx}, \quad (8.6.10)$$

will be periodic with the crystal lattice, which is fine, since the wave function doesn’t need to be periodic to satisfy Bloch’s theorem.

The central equation couples C_k to C_{k+nG} for $n \in \mathbb{Z}$. We can write the wave function as

$$\psi_k(x) = \sum_G C_{k-G} e^{i(k-G)x}, \quad (8.6.11)$$

for some fixed wave vector k . Rearranging this we have

$$\psi_k(x) = e^{ikx} \sum_G e^{-iGx} = e^{ikx} u_{k(x)} \quad (8.6.12)$$

here we identify $u_k(x)$ as the sum and we know that this has the periodicity of the lattice since we can also identify it as the Fourier series of some function, u_k . Hence $\psi_k(x) = u_k(x)e^{ikx}$ is of the form expected by Bloch’s theorem.

8.7 Fermi Surfaces

The **Fermi surface** is a surface in reciprocal space which is the boundary occupied and unoccupied states. We can think of it as the surface of constant energy with $\varepsilon(\mathbf{k}) = \varepsilon_F$. Fermi surfaces are important since electronic properties depend on electron states near the Fermi surface, with near typically meaning within $k_B T$, since these are the only states between which electrons can be thermally excited.

We can express many important electronic properties as integrals over the Fermi surface, FS, for example the electrical conductivity, σ :

$$\sigma \propto \int_{\text{FS}} \frac{v_x^2(\mathbf{k})}{v(\mathbf{k})} \tau(\mathbf{k}) d\mathbf{k}_{\text{FS}} \quad (8.7.1)$$

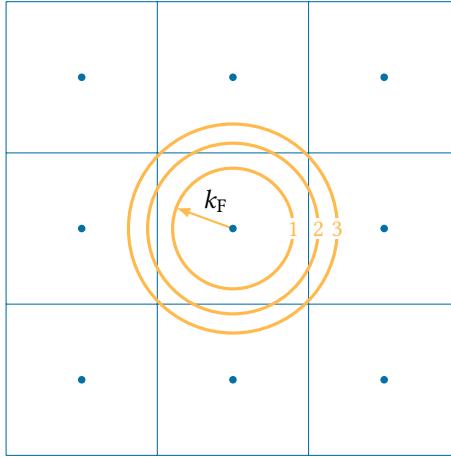


Figure 8.8: The Fermi surfaces for a square unit cell are circles, for divalent and trivalent crystals they extend beyond the first Brillouin zone.

where v is the electron velocity, τ the scattering time, and we assume an electric field in the x -direction. The electronic density of states can similarly be expressed as a Fermi surface integral, and the electronic density of states appears in many more properties, such as the electronic heat capacity,

$$C_{\text{el}} = \frac{1}{3} \pi^2 D(\varepsilon_F) k_B^2 T. \quad (8.7.2)$$

So far we have thought of the Fermi surface as a sphere, of radius k_F . This is valid for monovalent metals. For polyvalent metals the Fermi surface is more complicated. In particular it starts to extend beyond the first Brillouin zone.

Consider the example of a square unit cell in two dimensions with lattice parameter a . In a crystal with N unit cells there are $2N$ states in the square Brillouin zone, two for each allowed value of k , of which there are N .

For a monovalent crystal there is one valence electron per atom and there are N electrons in the crystal, hence

$$\pi k_F^2 = \frac{1}{2} \left(\frac{2\pi}{a} \right)^2 \implies k_F = \sqrt{\frac{2}{\pi}} \frac{\pi}{a}. \quad (8.7.3)$$

The left hand side here being the area contained by the Fermi surface (which is a circle in the two-dimensional, square unit cell case). The right hand side is half of the area of the unit cell in reciprocal space, since the unit cell contains $2N$ electron states, and we have N electrons.

For a divalent crystal we have

$$\pi k_F^2 = \left(\frac{2\pi}{a} \right)^2 \implies k_F = \sqrt{\frac{4}{\pi}} \frac{\pi}{a}. \quad (8.7.4)$$

For a trivalent crystal we have

$$\pi k_F^2 = \frac{3}{2} \left(\frac{2\pi}{a} \right)^2 \implies k_F = \sqrt{\frac{6}{\pi}} \frac{\pi}{a}. \quad (8.7.5)$$

In reciprocal space the Brillouin zone has side length $2\pi/a$, and so any value of k_F greater than π/a is going to extend beyond the first Brillouin zone. For the monovalent case $\sqrt{2/\pi} \approx 0.4$, so the Fermi surface is entirely within the first Brillouin zone. For the divalent case $\sqrt{4/\pi} \approx 0.56$, so the Fermi surface extends outside the first Brillouin zone.

We have seen that every wave vector outside of the first Brillouin zone is equivalent to another wave vector inside the first Brillouin zone. This means we can map the Fermi surface outside the first Brillouin zone back into the first Brillouin zone. This is what makes the Fermi surfaces of polyvalent crystals more complicated.

8.8 Higher Order Brillouin Zones

So far in this course we have considered only the first Brillouin zone, which we have simply been calling *the* Brillouin zone much of the time. Recall how we constructed this, using the Wigner–Seitz method, see [Section 6.3](#) for details. Essentially we create lines between lattice points, take the perpendicular bisectors, and then the smallest volume these enclose gives us our cell. In doing so we end up with bisectors that don't bound the first cell, instead we can use these to make a slightly larger, no longer primitive cell. We call the area in this cell, but not the first Brillouin zone, the second Brillouin zone. We can then do something similar to define the third Brillouin zone and so on. This is demonstrated in [Figure 8.10](#).

It is often easier to consider the alternative definition of the n th Brillouin zone, that a point, \mathbf{k} , is in the n th Brillouin zone if the condition $|\mathbf{k}| < |\mathbf{k} - \mathbf{G}|$ fails for exactly $n - 1$ reciprocal lattice vectors, \mathbf{G} . That is that \mathbf{k} is closer to the origin than it is to all but $n - 1$ lattice points. This is equivalent to the perpendicular bisector definition since moving across a perpendicular bisector means changing which of the two lattice points the point is closer to.

Each higher order Brillouin zone covers the same wave vectors as the first Brillouin zone, just shifted by some reciprocal lattice vector. This means that each Brillouin zone has the same volume. We can see this by imagining cutting up the Brillouin zones to fit in the first Brillouin zone. See [Figure 8.11](#).

Brillouin zones quickly become rather complicated shapes, even for simple lattices, like the square lattice.

8.9 Fermi Surfaces in the First Brillouin Zone

We can reduce a Fermi surface that intersects higher order Brillouin zones to a surface entirely in the first Brillouin zone. To do so we draw all Fermi surfaces, for each lattice point. We then consider only the Fermi surface inside the first Brillouin zone for some given lattice point. See [Figure 8.12](#).

Having done this we can then separate the reduced Fermi surface into parts, often called sheets, according to the Brillouin zone that they were in before mapping. Doing so we see that none of the Fermi surface comes from the first Brillouin zone, since the circle is entirely outside of the first Brillouin zone. The distorted square around the lattice point comes from the second Brillouin zone. The third Brillouin zone gives four pointed stars, one point of which is in the first Brillouin zone for each corner. The fourth Brillouin zone gives distorted squares on each corner, one quarter of which is in the first Brillouin zone.

All of this analysis can be applied to other lattices, and in higher dimensions. The Brillouin zones, and hence resulting Fermi surfaces, quickly become very complex.

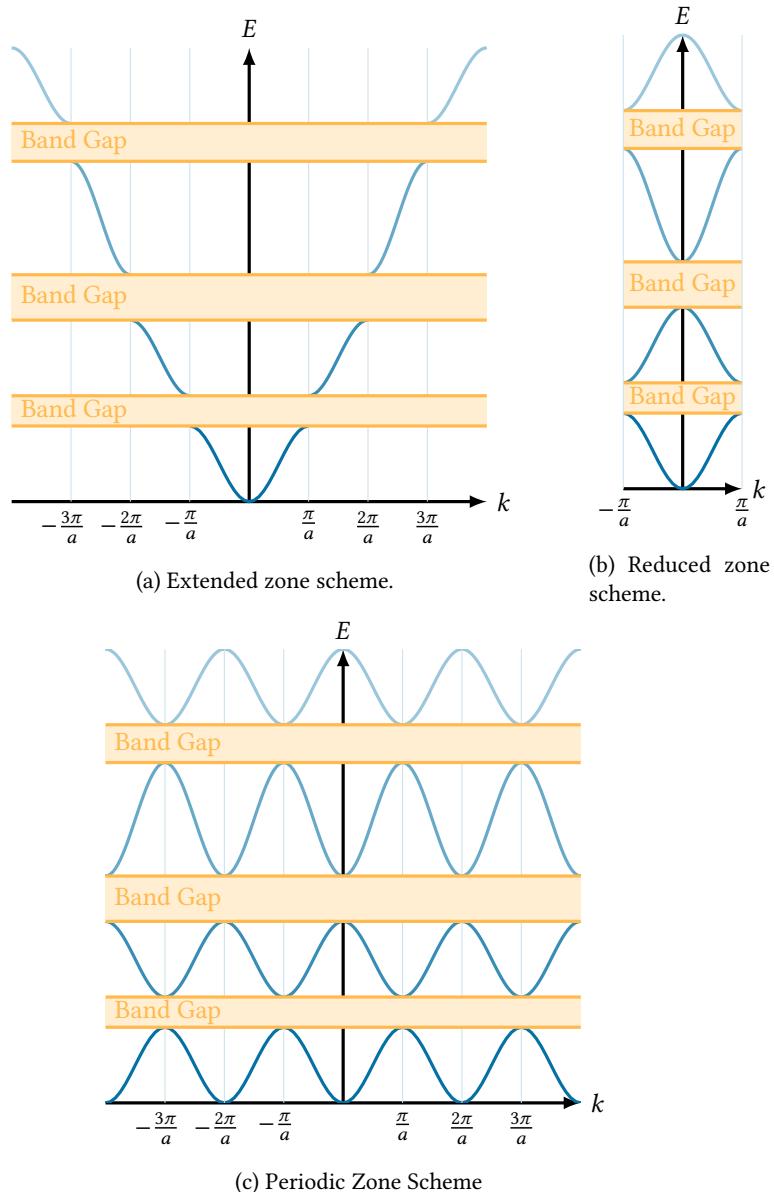


Figure 8.9: Different ways of viewing dispersion relations. In one dimension we can map an extended zone scheme into the first Brillouin zone by shifting by suitable multiples of $2\pi/a$, giving the reduced zone scheme. We can then extend this to all space by extending it periodically.

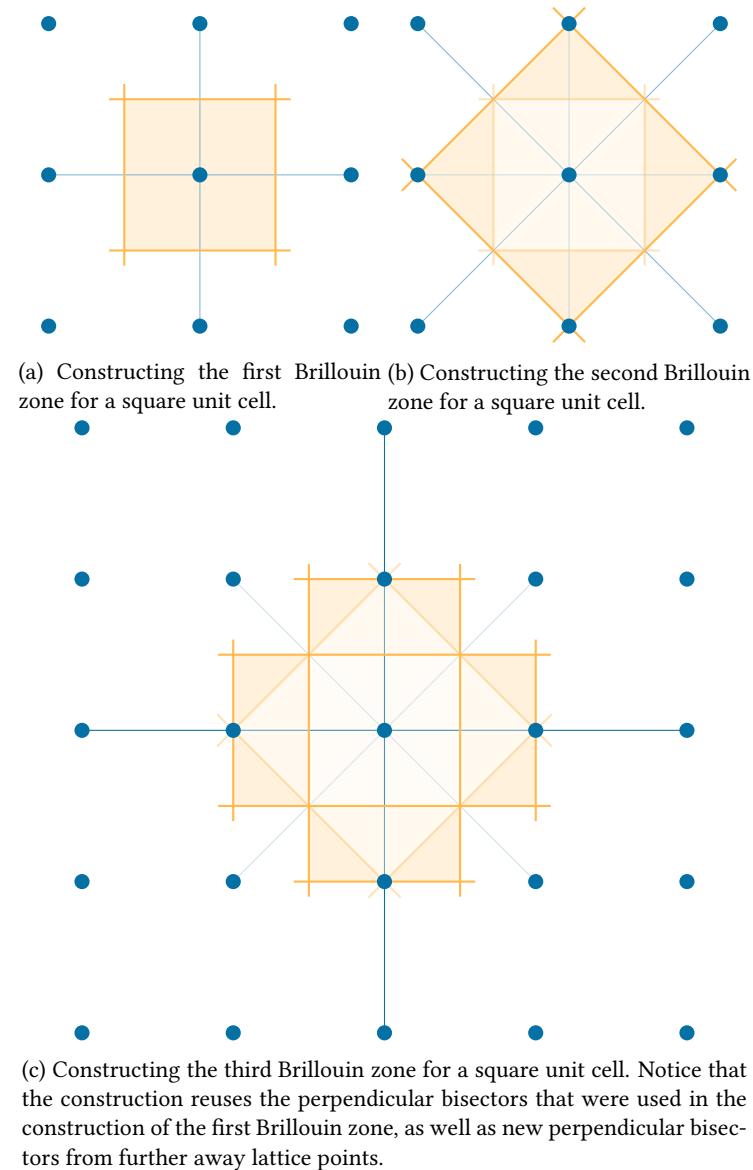
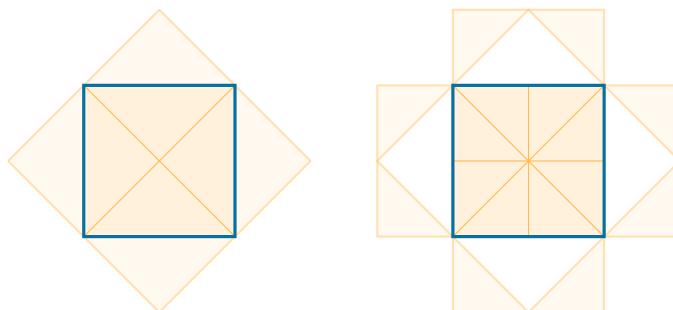


Figure 8.10: Extending the notion of Brillouin zones to higher orders.



(a) The second Brillouin zone has the same area as the first.
 (b) The third Brillouin zone has the same area as the first.

Figure 8.11: All Brillouin zones cover the same states, therefore they all have the same area.

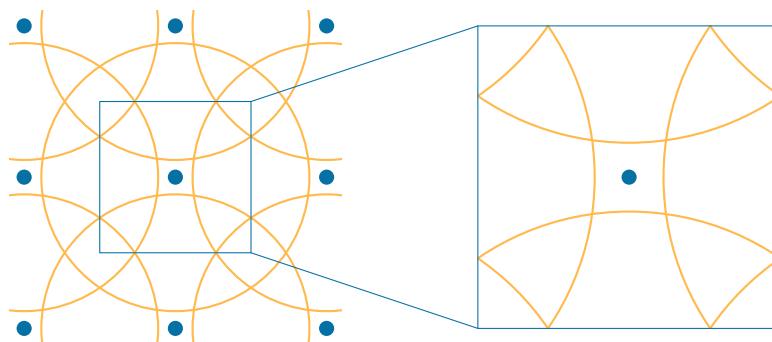


Figure 8.12: Mapping the Fermi surface to the first Brillouin zone by repeating the Fermi surface with the lattice and then ignoring everything outside of the first Brillouin zone for some given lattice point.

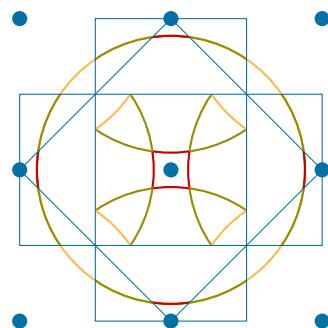


Figure 8.13: Colour coded reduced Fermi surface based on which Brillouin zone the Fermi surface was mapped from.

8.10 Sharp Edges

The Fermi surfaces that we have been drawing have sharp edges. This is not the case in reality. Modifications to the electron bands due to the actual crystal potential will give much more rounded edges. In fact it turns out that real Fermi surfaces will cross Bragg planes, which are the perpendicular planes to the lattice vectors, and hence planes parallel to the edges of the Brillouin zones, at right angles.

Physically we can argue that this is reasonable as follows. The group velocity of an electron with wave vector \mathbf{k} is given by $v_g = \nabla_{\mathbf{k}}\varepsilon(\mathbf{k})/\hbar$ where $\varepsilon(\mathbf{k})$ is the energy dispersion. We also know that, by definition, the Fermi surface is a surface of constant energy, $\varepsilon(\mathbf{k}) = \varepsilon_F$. Hence basic multivariable calculus tells us that $v_g \propto \nabla_{\mathbf{k}}\varepsilon$ must be perpendicular to the Fermi surface. If there is a sharp edge to the Fermi surface then the gradient is undefined at that point and hence the group velocity of the electron is undefined. This is not physically reasonable and so we conclude that these sharp edges don't reflect reality. Due to the way that the mapping to the first Brillouin zone occurs we see that if the Fermi surface always cuts through the boundaries between Brillouin zones at right angles then the resulting Fermi surface will have no sharp corners after we map it to the first Brillouin zone.

For our purposes we can construct a Fermi surface with sharp corners and simply round off the corners by making the surface cross the Brillouin zone boundary at right angles.

Nine

Semiconductors

9.1 Band Structure

Recall that for a three dimensional crystal with $N = N_x N_y N_z$ primitive unit cells with lattice parameters $L_i = N_i a_i$ the allowed wave vectors have components

$$k_i = 0, \pm \frac{2\pi}{L_i}, \pm \frac{4\pi}{L_i}, \dots, \frac{N_i \pi}{L_i}. \quad (9.1.1)$$

There are therefore N distinct wave vectors in the first Brillouin zone, since all states correspond uniquely to a wave vector in the first Brillouin zone. Allowing for electron spins there are $2N$ states in the first Brillouin zone.

Consider the model band structure shown in [Figure 9.1](#). Depending on the number of electrons per unit cell we get different types of material. A monovalent crystal will half fill the first energy band. This means that states near the Fermi surface can easily be excited into free states above the Fermi surface. This is a metal. A divalent crystal with this band structure will completely fill the first energy band. It is then not possible to excite electrons across the band gap at reasonable energies. This means that an electric field cannot accelerate electrons, hence the material an insulator. A trivalent crystal will partially fill the next energy band, so once again it is possible to excite electrons and it will be a metal again. See [Figure 9.2](#).

Not all divalent crystals are insulators. If there is band overlap then instead of filling the first energy band the next band up will be partially filled and the result is

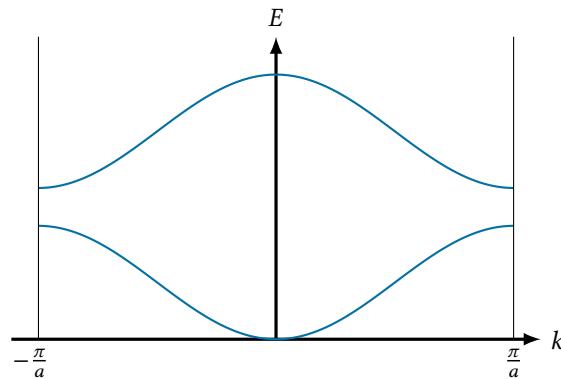
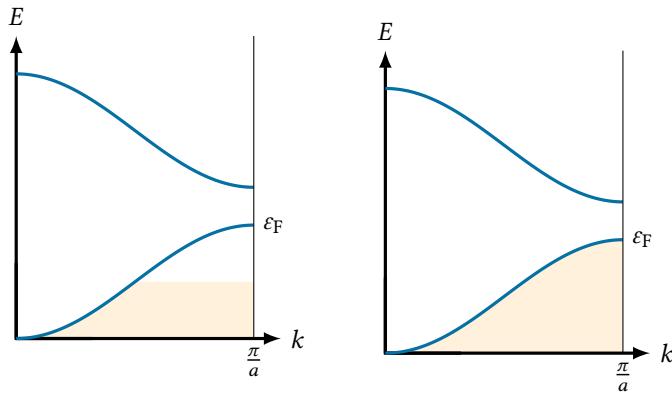
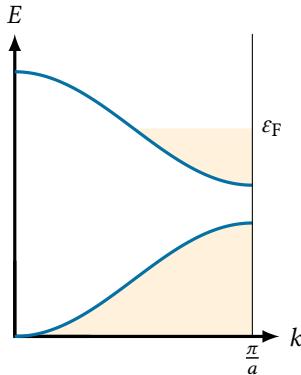


Figure 9.1: Model band structure.



(a) A metal that partially fills the first electron band. (b) An insulator fills up to a band gap.



(c) A metal that partially fills the second electron band.

Figure 9.2: The same band structure can give metals or insulators, depending on the number of electrons. Since the band structure is symmetric about the origin only positive wave vectors are shown.

that the material will have metallic properties. We call such a material a **semimetal**. See [Figure 9.3](#).

9.2 Semiconductors

A **semiconductor** is essentially an insulator where the band gap is small enough that a small number of electrons can be excited across the band gap. This allows some conductivity, but not as much as a metal. Clearly this is temperature dependent since a large amount of the required energy comes from thermal energy. What this means is that at sufficiently high temperatures all insulators will become semiconductors. This is what we meant by reasonable energy earlier, it's also likely that many insulators will melt before becoming semiconductors. On the other hand semiconductors become insulators if they are cooled too much.

There is no agreed upon size of band gap that makes something a semiconductor or an insulator. Typically insulators have a band gap greater than about 4 eV to 5 eV, for example diamond has a band gap of 5.5 eV and is definitely an insulator. Semiconductors then have smaller band gaps of a couple of electron volts.

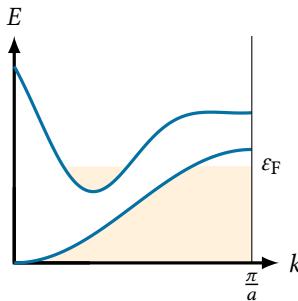


Figure 9.3: The band structure for a semimetal has overlapping bands, so the second band starts to fill up before the first band has finished filling up.

9.3 Direct and Indirect Band Gaps

We can broadly classify semiconductors into two categories, direct and indirect band gaps. A **direct band gap** has the maxima of the lower band at the same wave vector as the minima of the upper band. An **indirect band gap** doesn't. See [Figure 9.4](#).

We measure band gaps by measuring the absorbance of light. In general if the intensity of light incident on the sample is I_{in} and the intensity of the light passes through the sample is I_{out} then we define the **transmittance** as $T := I_{\text{out}}/I_{\text{in}}$. We then define the **absorbance** as

$$A := \log_{10} \left(\frac{I_{\text{in}}}{I_{\text{out}}} \right) = \log_{10} \frac{1}{T} = -\log_{10} T. \quad (9.3.1)$$

The absorbance depends on the sample thickness. The Beer–Lambert law gives the absorbance as a function of the sample thickness, L :

$$I_{\text{out}} = I_{\text{in}} e^{-\alpha L} \quad (9.3.2)$$

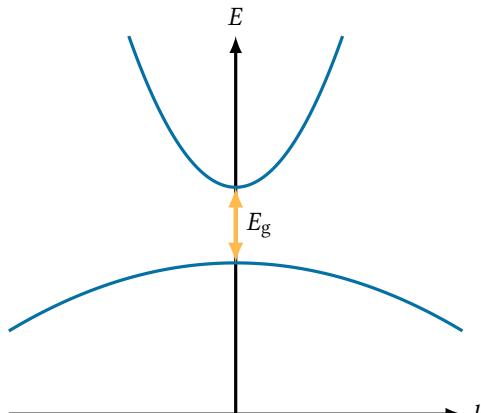
where α is the **absorption coefficient** defined as

$$\alpha = \frac{A}{L} \ln 10. \quad (9.3.3)$$

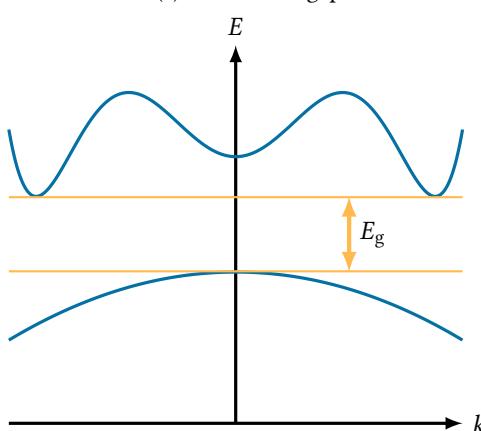
Absorbance, A , is dimensionless and the absorption coefficient, α , has dimensions of inverse length.

The wave vector of optical photons is negligible compared to the wave vector at the Brillouin zone boundary. Therefore the wave vector of an electron doesn't change noticeably when a photon is absorbed. This means that the electron has to be excited to a state with (approximately) the same wave vector, which means straight upward in the band scheme.

The process for measuring whether a band gap is direct or indirect involves measuring the absorbance for a range of photon frequencies. For a direct band gap the absorbance is zero below ω_g , where ω_g is such that $E_g = \hbar\omega_g$. For an indirect band gap on the other hand there will be a small number of absorption processes where a phonon, of frequency Ω , is annihilated at the same time, donating its wave vector, and energy, $\hbar\Omega$, to the electron in addition to the energy, $\hbar\omega$, donated by the photon. This means that a small number of events occur where electrons are excited when $\omega < \omega_g$. This is a relatively rare occurrence due to the need for three particles



(a) Direct band gap.



(b) Indirect band gap.

Figure 9.4: Direct and indirect band gaps.

to interact. The resulting shape of a plot of absorbance against photon energy has a characteristic lead in which allows us to distinguish between direct and indirect band gaps. See Figure 9.5.

9.4 Wave Vectors

Recall that the group velocity is $v_g = \nabla_{\mathbf{k}}\varepsilon(\mathbf{k})/\hbar$. The work done, $\delta\varepsilon$, in an electric field during a time interval δt is

$$\delta\varepsilon = F\delta x = -eEv_g\delta t. \quad (9.4.1)$$

It is also

$$\delta\varepsilon = \frac{d\varepsilon}{dk}\delta k = \hbar v_g\delta k. \quad (9.4.2)$$

Combining these and moving to three dimensions

$$\hbar\delta k = -eE\delta t \implies \hbar \frac{d\mathbf{k}}{dt} = -e\mathbf{E} = \mathbf{F}. \quad (9.4.3)$$

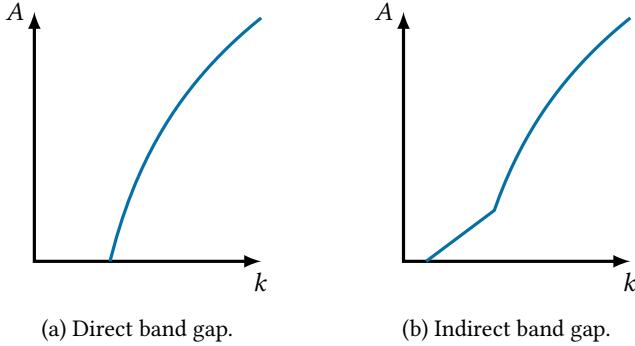


Figure 9.5: Absorbance of direct and indirect band gaps.

This is the force on an electron in a periodic potential, recall that e is positive so $-e$ is the electron charge. Putting this in the Lorentz force equation we have

$$\mathbf{F} = \hbar \frac{d\mathbf{k}}{dt} = -e(\mathbf{E} + \mathbf{v} \times \mathbf{B}) = -e(\mathbf{E} + \frac{1}{\hbar} \nabla_{\mathbf{k}} \epsilon \times \mathbf{B}). \quad (9.4.4)$$

9.5 Holes

Consider what happens when an electron is excited from the valence band to the conduction band. Suppose we start with 10^{20} electrons. We can then describe the properties of the valence band in terms of the $10^{20} - 1$ populated valence band states. Alternatively, we can describe them in terms of the one vacant state. Clearly this is much easier to do. We treat the vacant state as a quasiparticle, which we call a **hole**.

With no excited electrons the total wave vector of the filled valence band is 0. Removing an electron with wave vector \mathbf{k}_e means that the total wave vector of the remaining electrons in the valence band is $-\mathbf{k}_e$. We can identify this as the wave vector of the hole, $\mathbf{k}_h = -\mathbf{k}_e$.

Similarly the energy of the hole state is the missing energy of the removed electron, that is $\epsilon_h(\mathbf{k}_h) = -\epsilon_e(\mathbf{k}_e)$.

At zero kelvin we can place the Fermi energy, that is the highest point in the valence band, at zero. We then get that the flipped valence band gives the properties of the hole. See Figure 9.6.

To find the group velocity of the hole we can simply use the hole wave vector and dispersion relation, the identity $\nabla_{-\mathbf{k}} f(\mathbf{k}) = -\nabla_{\mathbf{k}} f(\mathbf{k})$, and the fact that, for non-magnetic materials, the dispersion relations are symmetric under inversion so $\epsilon_e(-\mathbf{k}_e) = \epsilon_e(\mathbf{k}_e)$:

$$\mathbf{v}_h(\mathbf{k}_h) = \nabla_{\mathbf{k}_h} \epsilon_h(\mathbf{k}_h) = \nabla_{-\mathbf{k}_e} (-\epsilon(-\mathbf{k}_e)) = \nabla_{\mathbf{k}_e} \epsilon(-\mathbf{k}_e) = \nabla_{\mathbf{k}_e} \epsilon(\mathbf{k}_e) = \mathbf{v}_e(\mathbf{k}_e). \quad (9.5.1)$$

So the velocity of the hole is the same as the velocity of the electron.

We will introduce the concept of effective mass, m^* , in a later section. For now we just talk of the mass and make a note here that what we are actually considering is the effective mass. Analogously to the hole wave vector and energy being the negative of the electrons since they represent “missing” quantities the hole mass is $m_h = -m_e$.

The charge of the hole is the missing charge of the removed electron, so $q_h = -q_e$.

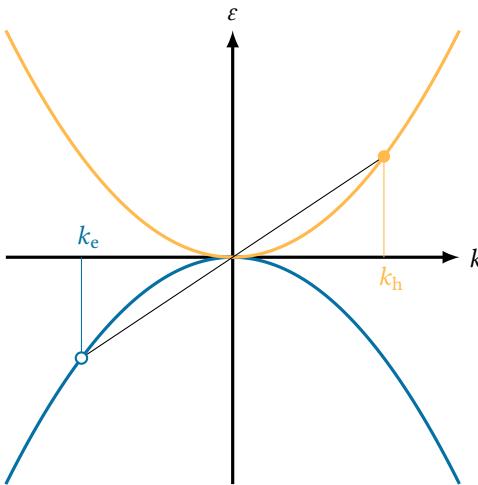


Figure 9.6: The hole band is constructed by setting $\mathbf{k}_h = \mathbf{k}_e$ and $\varepsilon_h(\mathbf{k}_h) = -\varepsilon_e(\mathbf{k}_e)$.

The equation of motion for the hole is then

$$\hbar \frac{d\mathbf{k}_h}{dt} = e(E + \mathbf{v}_h \times \mathbf{B}). \quad (9.5.2)$$

Since $\mathbf{v}_h = \mathbf{v}_e$ the only difference here is that for an electron there is an extra overall negative due to the charge. So we see that a hole behaves like a quasiparticle with positive charge, e .

Now consider what happens when we apply an electric field, say in the positive x -direction (by which we really mean the direction $(1, 0, 0)$ in reciprocal space). As long as there is no scattering the electrons, which are negative, are all accelerated in the negative x -direction. The state where there is a hole is filled in by the electron to its right and the hole therefore moves right, in the positive x -direction. The electron in the state at the left edge of the Brillouin zone, $k = -\pi/a$, moves to the equivalent state $k' = k + 2\pi/a$.

The hole moves in the opposite direction to the electrons and has opposite charge. Therefore the current is $\mathbf{j} = (-e)\mathbf{v}_{\text{hole}} = e\mathbf{v}_{\text{electron}}$. Note that the velocities here have *opposite signs* since we are considering the motion of a hole and a different electron. Before, when we concluded that the electron and hole velocities are equal, we were considering the hole and the electron that was removed to create it. So the current due to motion of an electron is the same as the current due to the motion of the hole. This means that we can describe the current from the nearly filled valence band in terms of either holes or electrons.

One place where the distinction between electrons and holes is important is the Hall effect. When we calculate the Hall coefficient for holes we will get the *negative* of the Hall coefficient for electrons. Since the Hall coefficient for electrons is negative this results in a positive Hall coefficient for holes. We can use this to measure whether holes or electrons dominate current flow. For example, beryllium is a semimetal with one nearly full band and one mostly empty band. The Hall effect from electrons and holes almost cancels but not quite and the resulting Hall coefficient is positive as the holes contribute slightly more to the current.

Table 9.1: Summary of hole properties.

Quantity	Electron	Hole
Wave Vector	\mathbf{k}_e	$\mathbf{k}_h = -\mathbf{k}_e$
Energy	$\varepsilon_e(\mathbf{k}_e)$	$\varepsilon_h(\mathbf{k}_h) = -\varepsilon_e(\mathbf{k}_e)$
Velocity	v_e	$v_h = -v_e$
Effective Mass	m_e^*	$m_h^* = -m_e^*$
Charge	q_e	$q_h = -q_e$
Current	j	j

9.6 Effective Mass

For the free electron dispersion relation,

$$\varepsilon(k) = \frac{\hbar^2}{2m} \quad (9.6.1)$$

the curvature of the band is determined, as usual, by the second derivative, which in this quadratic case is simply proportional to the coefficient, $\hbar^2/(2m)$. We see, therefore, that the band curvature depends on the electron mass. In this section we ask whether this relation is reversible, can we find the electron mass from the curvature? We can't, but we can find a quantity with dimensions of mass that has physical meaning.

Recall that we found

$$\hbar \frac{dk}{dt} = F \quad (9.6.2)$$

earlier in [Equation \(9.4.3\)](#), and that the group velocity is defined as

$$v_g = \frac{1}{\hbar} \frac{d\varepsilon}{dk} \quad (9.6.3)$$

The acceleration is then

$$\frac{dv_g}{dt} = \frac{1}{\hbar} \frac{d}{dt} \frac{d\varepsilon}{dk} = \frac{1}{\hbar} \frac{d^2\varepsilon}{dk^2} \frac{dk}{dt} = \frac{1}{\hbar^2} \frac{d^2\varepsilon}{dk^2} F. \quad (9.6.4)$$

Rearranging this we have

$$F = \hbar^2 \left(\frac{d^2\varepsilon}{dk^2} \right)^{-1} \frac{dv_g}{dt} = m^* \frac{dv_g}{dt}. \quad (9.6.5)$$

Here we have defined the **effective mass**

$$m^* := \hbar^2 \left(\frac{d^2\varepsilon}{dk^2} \right)^{-1}. \quad (9.6.6)$$

That is the effective mass is defined to be inversely proportional to the band curvature.

This definition of the effective mass coincides with the electron mass in the free electron case. In reality the effective mass and electron mass will be different. We can think of the effective mass as the electron mass corrected for the potential. It should be noticed that the effective mass can be negative. This corresponds to an electron being back scattered, being accelerated in the opposite direction to what one would expect for a free electron. Replacing the electron mass with the effective mass in many of our previous results will give a more accurate value.

9.6.1 In Three Dimensions

In three dimensions the curvature is, in general, a tensor quantity. Generalising the definition of the effective mass we define the **effective mass tensor**,

$$\left(\frac{1}{m^*}\right)_{ij} := \frac{1}{\hbar^2} \frac{d}{dk_i} \frac{d\varepsilon}{dk_j}. \quad (9.6.7)$$

This relates to the acceleration by

$$\frac{dv_i}{dt} = \sum_{j=1}^3 \left(\frac{1}{m^*}\right)_{ij} F_j. \quad (9.6.8)$$

9.7 Typical Semiconductor Band Structures

9.7.1 Silicon

Silicon has a diamond like structure, that means an FCC lattice, but not an FCC structure. It has an indirect band gap of $E_g = 1.17$ eV at 0 K, and $E_g = 1.12$ eV at 300 K. Both much larger than the energy, $k_B T \approx 25$ meV, available at room temperature. The lowest band is nearly parabolic and has a minimum at the Γ point ($\mathbf{k} = 0$). This is pretty much what we would expect from the free electron approximation. The maximum of the valence band also occurs near the Γ point. The minimum of the conduction band occurs near the X point, which makes the band gap indirect.

9.7.2 Germanium

Germanium has the diamond structure, similar to silicon. It has a similar band structure also. The band gap is smaller, $E_g = 0.75$ eV at 0 K, or $E_g = 0.67$ eV at 300 K. The lowest band and valence band both have extrema at the Γ point, like silicon. The minimum of the conduction band is at the L point. This is an indirect band gap. Spin-orbit coupling lifts degeneracies and creates band splitting compared to silicon.

9.7.3 Gallium Arsenide

Gallium Arsenide has the zinc blende structure, which is the diamond structure but with two types of atom. It has a direct band gap at the Γ point. The band gap is $E_g = 1.52$ eV at 0 K, and $E_g = 1.43$ eV at 300 K.

9.8 Intrinsic Semiconductors

Semiconductors are characterised by having electrical conductivity at room temperature which is several orders of magnitude lower than a metal, but not zero like a conductor. The electrical conductivity increases exponentially with temperature. It is this property that we will now look at, starting with intrinsic semiconductors. An **intrinsic semiconductor** is a perfect semiconductor without structural defects or chemical impurities. This means no doping. Clearly intrinsic semiconductors don't exist in real life but there are many materials that are reasonably close.

Current is carried by free charge carriers. In a semiconductor each excited electron gives *two* free charge carriers, the electron itself and the hole. Most of the excited electrons are thermally excited, but some will also be excited by absorption of a photon. At 0 K we expect that an intrinsic semiconductor will have zero conductivity.

We can express the conductivity, σ , in terms of the electron and hole concentrations, n and p respectively (for negative and positive), as well as the electron and hole mobilities, μ_n and μ_p . The charge carrier **mobility**, μ , is a quantity relating the carrier drift velocity, v_d , to the electric field, $v_d = \mu E$. For a free electron Fermi gas model the mobility is $\mu = e\tau/m^*$, where τ is the average time between scattering events, e is the elementary charge, and m^* is the effective mass. In this context we often call the free electron Fermi gas model the **Drude–Sommerfeld model**.

9.8.1 Conductivity

With the starting point of the free electron gas model the conductivity is

$$\sigma = \frac{ne^2\tau}{m^*}. \quad (9.8.1)$$

Including holes and writing this in terms of the mobility the conductivity is

$$\sigma = |e|(n\mu_n + p\mu_p). \quad (9.8.2)$$

Note that both electrons and holes contribute with the same sign, since the current for holes is the same as the current for electrons in the same magnetic field as the direction of travel and charge are both reversed, cancelling out.

The temperature dependence comes from the temperature dependence of n and p . In order to predict this we model the valence and conduction bands as one parabolic band each, more accurate band structures don't yield analytic results and we have to resort to numerical methods. We then make use of results from the free electron gas model, in particular the density of states.

Annoyingly there are two different ways that the density of states is defined. So far we have been using $D(\varepsilon)$ taking

$$\int_0^\infty D(\varepsilon)f(\varepsilon, T) d\varepsilon \quad (9.8.3)$$

to give the number of electrons. In this section we will use the density of states, $\tilde{D}(\varepsilon)$, such that

$$\int_0^\infty \tilde{D}(\varepsilon)f(\varepsilon, T) d\varepsilon \quad (9.8.4)$$

gives the *density* of electrons. So $\tilde{D} = D/V$. Both D and \tilde{D} are usually denoted D and it is left to context to determine which is being used. It is this second type, \tilde{D} , which we make use of in this section, and following conventions, we denote it D from here on.

The electron concentration is

$$n = \int_{\varepsilon_c}^\infty D_c(\varepsilon)f(\varepsilon, T) d\varepsilon \quad (9.8.5)$$

where ε_c is the minimum energy in the conduction band. For the conduction band the density of states is

$$D_c(\varepsilon) = \frac{(2m_n^*)^{3/2}}{2\pi^2\hbar^3} \sqrt{\varepsilon - \varepsilon_c}, \quad \text{for } \varepsilon > \varepsilon_c. \quad (9.8.6)$$

At normal temperatures, such that $k_B T \ll \varepsilon - \varepsilon_F$, we can approximate the Fermi–Dirac distribution using the fact that $e^{\beta(\varepsilon - \varepsilon_F)}$ will be small:

$$f(\varepsilon) - \frac{1}{e^{\beta(\varepsilon - \varepsilon_F)} + 1} \approx e^{-\beta(\varepsilon - \varepsilon_F)} = e^{\beta\varepsilon_F} e^{-\beta\varepsilon}. \quad (9.8.7)$$

We then have

$$n = \int_{\varepsilon_c}^{\infty} D_c(\varepsilon) f(\varepsilon, T) d\varepsilon \quad (9.8.8)$$

$$\approx \frac{(2m_n^*)^{3/2}}{2\pi^2\hbar^3} e^{\beta\varepsilon_F} \int_{\varepsilon_c}^{\infty} e^{-\beta\varepsilon} \sqrt{\varepsilon - \varepsilon_c} d\varepsilon. \quad (9.8.9)$$

Considering this integral, defining $x = \beta(\varepsilon - \varepsilon_c)$ we have

$$n \approx \frac{(2m_n^*)^{3/2}}{2\pi^2\hbar^3} (k_B T)^{3/2} e^{-\beta(\varepsilon_c - \varepsilon_F)} \int_0^{\infty} x^{1/2} e^{-x} dx \quad (9.8.10)$$

$$= \frac{(2m_n^*)^{3/2}}{2\pi^2\hbar^3} (k_B T)^{3/2} e^{-\beta(\varepsilon_c - \varepsilon_F)} \Gamma\left(\frac{3}{2}\right) \quad (9.8.11)$$

$$= \frac{(2m_n^*)^{3/2}}{2\pi^2\hbar^3} (k_B T)^{3/2} e^{-\beta(\varepsilon_c - \varepsilon_F)} \frac{\sqrt{\pi}}{2} \quad (9.8.12)$$

where we have recognised the gamma function

$$\Gamma(z+1) := \int_0^{\infty} \int_0^{\infty} x^z e^{-x} dx, \quad \text{and} \quad \Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2}. \quad (9.8.13)$$

This gives us the final result

$$n(T) = 2 \left(\frac{m_n^* k_B T}{2\pi\hbar^2} \right)^{3/2} e^{-\beta(\varepsilon_c - \varepsilon_F)}. \quad (9.8.14)$$

We can make an analogous computation for the holes in the valence band, the only difference being we are now considering hole energies below the Fermi level so $-(\varepsilon_c - \varepsilon_F) \rightarrow -(\varepsilon_F - \varepsilon_v) = \varepsilon_v - \varepsilon_F$. The hole density is then

$$p(T) = 2 \left(\frac{m_p^* k_B T}{2\pi\hbar^2} \right)^{3/2} e^{\beta(\varepsilon_v - \varepsilon_F)}. \quad (9.8.15)$$

In an intrinsic semiconductor every free electron has a corresponding free hole. Therefore the number of electrons and holes, and hence their concentrations, are the same. We can use this to eliminate the unknown Fermi level, ε_F . We also add subscript i to denote that this only applies to intrinsic semiconductors. Doing so we have

$$n_i = p_i = 2 \left(\frac{k_B T}{2\pi\hbar^2} \right)^{3/2} (m_n^* m_p^*)^{3/2} e^{-E_g \beta/2}. \quad (9.8.16)$$

Here we have identified $\varepsilon_c - \varepsilon_v = E_g$.

From this we see that the temperature dependence is mostly from the $e^{-E_g/(2k_B T)}$ term. There is also a small correction from the $T^{3/2}$ term. As well as the temperature dependence here the electron and hole mobilities are also temperature dependent, although this effect is much smaller than the exponential dependence in n_i and p_i . This matches experimental data quite well if we choose the effective masses m_n^* and m_p^* as parameters to our model. To get a sense of size of these quantities see Table 9.2.

Table 9.2: Typical band gaps and free electron concentrations for some intrinsic semiconductors and the free electron concentrations of some metals for comparison.

Material	E_g/eV	n_i/cm^{-3}
Ge	0.67	2.4×10^{13}
Si	1.1	1.5×10^{10}
GaAs	1.43	5.0×10^7
Na	N/A	2.7×10^{22}
Al	N/A	18×10^{22}

9.9 Temperature Dependence of the Fermi Level

An intrinsic neutral semiconductor must be charge neutral at all temperatures. This means that $n_i(T) = p_i(T)$. The density of occupied conduction band states is given by

$$\frac{dn}{d\varepsilon} = D_c(\varepsilon)f(\varepsilon, \varepsilon_F, T). \quad (9.9.1)$$

For the valence band holes we need to consider the missing electrons, so we replace $f(\varepsilon, \varepsilon_F, T)$ with $1 - \varphi(\varepsilon, \varepsilon_F, T)$. Integrating the occupied densities of states over the relevant energy ranges for the conduction and valence band we get the electron and hole concentrations. That is the concentrations are given by the area under the $dn/d\varepsilon$ and $dp/d\varepsilon$ curves.

Consider the case of $m_p^* = 2m_n^*$, and ε_F is directly between ε_v and ε_c . In this case p_i will be larger than n_i . Since this can't be true we see that the Fermi level must be at a different point. Shifting it up slightly we will find a value such that $n_i = p_i$.

In general by comparing the densities we have

$$n_i(T) = p_i(T) \quad (9.9.2)$$

$$(m_n^*)^{3/2} e^{-\beta(\varepsilon_c - \varepsilon_F)} = (m_p^*)^{3/2} e^{\beta(\varepsilon_v - \varepsilon_F)} \quad (9.9.3)$$

$$\frac{e^{-\beta(\varepsilon_c - \varepsilon_F)}}{e^{\beta(\varepsilon_v - \varepsilon_F)}} = \frac{(m_p^*)^{3/2}}{(m_n^*)^{3/2}} \quad (9.9.4)$$

$$[(\varepsilon_F - \varepsilon_c) - (\varepsilon_v - \varepsilon_F)] = \frac{3}{2} k_B T \ln \left(\frac{m_p^*}{m_n^*} \right) \quad (9.9.5)$$

$$\varepsilon_F = \frac{1}{2}(\varepsilon_c + \varepsilon_v) + \frac{3}{4} k_B T \ln \left(\frac{m_p^*}{m_n^*} \right). \quad (9.9.6)$$

This shows that at 0 K the Fermi level is in the middle of the band gap and then at higher temperatures shifts by an amount proportional to $k_B T$. The sign of the shift depends on the ratio of the effective masses. The value of ε_F depends on the density of states, through m_n^* and m_p^* .

For example, consider silicon. The band gap is $E_g = 1.1 \text{ eV}$ at 300 K. The electron mass is $m_n^* = 1.1 m_e$, and the hole mass is $m_p^* = 0.7 m_e$. At temperature T the shift from the 0 K value of ε_F is

$$\frac{3}{4} k_B T \ln \left(\frac{0.7}{1.1} \right) \approx -0.45 k_B T. \quad (9.9.7)$$

For example when going from 0 K to 300 K the Fermi level shifts by approximately $-0.45 \cdot 25 \text{ meV} \approx -11 \text{ meV}$. In general the shift in Fermi level will be small compared to the size of the band gap.

9.10 Doping

In a real material we cannot avoid impurities. These can greatly change the properties of the material. For example, a high purity sample of GaAs has a carrier concentration of approximately 10^{16} cm^{-3} . On the other hand for a truly pure sample we expect the intrinsic density to be approximately $5 \times 10^7 \text{ cm}^{-3}$. Another example is silicon, which has an intrinsic carrier concentration of approximately $1.5 \times 10^{10} \text{ cm}^{-3}$. This is not large enough to be useful in semiconductor devices.

We modify carrier charge density by intentionally introducing chemical impurities, a process called **doping**. Typically there are two types of doping materials. Donors give electrons to the crystal and trap holes, this increases n and decreases p . We call the resulting material *n*-type. Acceptors trap electrons and release holes, this increases p and decreases n . We call the resulting material *p*-type.

Donors have more valence electrons than are needed for bonding in the material being doped, and acceptors have fewer. For example Si and Ge are tetravalent, that is each atom forms four bonds. In this context P, As, and Sb are donors as they are pentavalent, so adding some of these adds an equal amount of weakly electrons. B, Al, Ga, and In are trivalent, and so acceptors in this context. Adding some of these adds weakly bound holes.

9.11 Charge Carrier Density

Recall that the charge carrier densities are

$$n(T) = 2 \left(\frac{m_n^* k_B T}{2\pi\hbar^2} \right)^{3/2} e^{-\beta(\varepsilon_c - \varepsilon_F)}, p(T) = 2 \left(\frac{m_p^* k_B T}{2\pi\hbar^2} \right)^{3/2} e^{\beta(\varepsilon_v - \varepsilon_F)}. \quad (9.11.1)$$

The derivation up to this point works for doped and intrinsic semiconductors. The analysis breaks down when we set $n_i = p_i$, which holds only for intrinsic semiconductors. The assumptions made to derive this model was that the bands are approximately parabolic and that $\varepsilon_c - \varepsilon_F, \varepsilon_F - \varepsilon_v \gg k_B T$. Consider the product

$$np = 4 \left(\frac{k_B T}{2\pi\hbar^2} \right)^3 (m_n^* m_p^*)^{3/2} e^{-\beta E_g}. \quad (9.11.2)$$

This is the same result as we had for $n_i p_i$, the only thing that changes between an intrinsic semiconductor and a doped semiconductor is the balance of electrons and holes. In an intrinsic semiconductor we have equal amounts of both whereas in a doped semiconductor we won't, but their product will be the same. This is known as the **law of mass action**.

9.12 Impurity as Hydrogen Like Centre

We will consider here a donor impurity but the analysis applies to acceptor impurities also. The donor atom will have some excess positive charge, compared to the cores of the other atoms in the material. The weakly bound electron can be treated like

the electron in a Hydrogen atom, with a different nuclear charge. In particular recall that the Rydberg energy levels are given by

$$E_n^H = \frac{m_e e^4}{2(4\pi\epsilon_0\hbar)^2} \frac{1}{n^2}. \quad (9.12.1)$$

For the impurity we replace m_e with m_n^* , and ϵ_0 with $\epsilon_0\epsilon$, where ϵ is the dielectric constant for the crystal. We then have

$$E_n^{\text{donor}} = \frac{m_n^* e^4}{2(4\pi\epsilon_0\epsilon\hbar)^2} \frac{1}{n^2} = \frac{m_n^*/m_e}{\epsilon^2} E_n^H. \quad (9.12.2)$$

For example, in silicon $m_n^* \approx 0.3 m_e$, $\epsilon \approx 11.7$, so

$$\frac{m_n^*/m_e}{\epsilon^2} \approx 0.0022. \quad (9.12.3)$$

Therefore, recalling that $E_1^H = 13.6 \text{ eV}$, the ground state binding energy for donor doping in silicon is

$$E_1^{\text{donor}} \approx 0.0022 \cdot 13.6 \text{ eV} \approx 30 \text{ meV}. \quad (9.12.4)$$

We see that the donor electrons are only weakly bound compared to the band gap, which is about 1.1 eV for silicon. This means that the donor electron is easily liberated to become a charge carrier, which is exactly what we want. In particular at room temperature $k_B T \approx 25 \text{ meV}$ so a lot of donor electrons will be released purely by thermal fluctuations.

We can define a quantity analogous to the Bohr radius in hydrogen:

$$r = \epsilon_0 \epsilon \frac{\hbar^2}{\pi m_n^* e^2} = \epsilon \frac{m_e}{m_n^*} a_0. \quad (9.12.5)$$

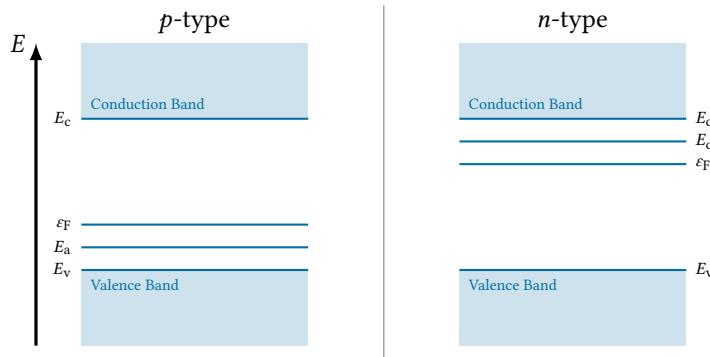
This can be thought of as a measure of how tightly bound the electron is to the donor atom. Going back to the donor doped silicon example we have that $r \approx (11.7/0.3)a_0 \approx 20 \text{ \AA}$. In a standard silicon crystal the interatomic distance is 2.35 \AA. From this we see that the donor electron cloud, which we take to be the volume enclosed by a sphere of radius r centred on the donor atom, extends over approximately 1000 silicon atoms. This justifies the need for the factor of ϵ since the donor electron is screened from the donor nucleus by many silicon atoms, which we account for with this factor of ϵ .

Further verification of this model is provided by accurately predicting absorbance patterns.

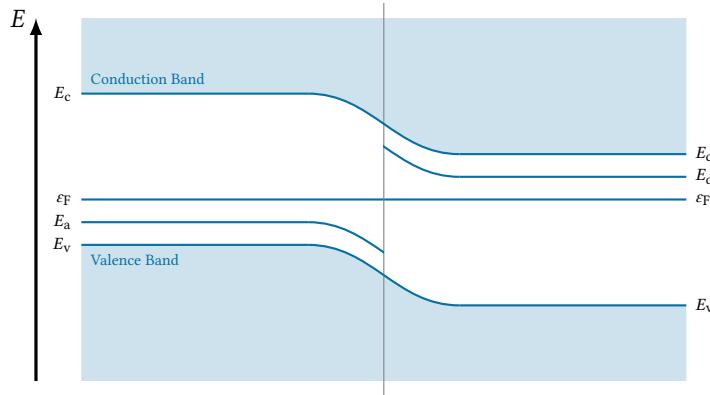
For an acceptor impurity instead of energy levels just below the valence band the hole has energy levels just above the conduction band.

9.13 Semiconductor Devices

A semiconductor device is usually formed by placing some *p*-type material next to some *n*-type material. This produces an interface, known as a *p-n* junction. Bringing these two materials together we get a boundary region where the energy levels and bands move smoothly from one to the other. The Fermi level is the same for the entire crystal, including across the junction. See Figure 9.7. Depending on the desired result



(a) The bands and energy levels of the doped electrons in *p* and *n*-type semiconductors.



(b) In a *p-n* junction the Fermi level is constant across the crystal and the other energy levels and bands are shifted to make this true. This means that a conduction electron needs to increase in energy to move from the *n*-type to *p*-type. This increase is on the order of the band gap, so about 1 eV

Figure 9.7: Forming a *p-n* junction from *p*-type and *n*-type doped semiconductors.

we can then apply a current to this either from *p* to *n* or *n* to *p*. If we do so one way then we will add electrons to the *n*-type increasing the difference between the two sides of the junction and allowing current flow. If we do so the other way no current can pass, this is the basics of how a diode is formed.

Many other devices can be formed in a similar way, such as transistors, the simplest of which have *npn* junctions and *pnp* junctions. Construction of semiconductor devices is a whole course to itself so we shall stop here.

Appendices

A

Constants and Useful Values

Table A.1: Important Constants and Their Values. Note masses in electron volts have implicit factors of c .

Constant	Symbol	Value
Speed of Light	c	$2.997\ 924\ 58 \times 10^8 \text{ m s}^{-1}$
Planck Constant	h	$6.626\ 070\ 15 \times 10^{-34} \text{ J s}$ $4.135\ 667\ 696 \text{ eV s}$
Reduced Planck Constant	\hbar	$1.054\ 571\ 817 \times 10^{-34} \text{ J s}$ $6.582\ 119\ 569 \times 10^{-16} \text{ eV s}$
Electron Mass	m_e	$9.109\ 383\ 701\ 5 \times 10^{-31} \text{ kg}$ $510.998\ 95 \text{ keV}$
Elementary Charge	e	$1.602\ 176\ 634 \times 10^{-19} \text{ C}$ $4.803\ 204\ 25 \times 10^{-10} \text{ statC}$
Electric Constant	ϵ_0	$8.854\ 187\ 812\ 8 \times 10^{-12} \text{ F m}^{-1}$
Boltzmann Constant	k_B	$1.380\ 649 \times 10^{-23} \text{ J K}^{-1}$ $8.617\ 333\ 262 \times 10^{-5} \text{ eV K}^{-1}$
Molar Gas Constant	R	$8.314\ 462\ 618 \text{ J mol}^{-1} \text{ K}^{-1}$
Avogadro Constant	N_A	$6.022\ 140\ 76 \times 10^{23} \text{ mol}^{-1}$

Table A.2: Some unit conversions. Note masses in electron volts have implicit factors of c .

$$\begin{array}{ll}
 1 \text{ kg} = 5.609 \times 10^{29} \text{ MeV} & 1 \text{ MeV} = 1.783 \times 10^{-30} \text{ kg} \\
 1 \text{ kg} = 6.022 \times 10^{26} \text{ Da} = 6.022 \times 10^{26} \text{ u} & 1 \text{ Da} = 1 \text{ u} 1.661 \times 10^{-27} \text{ kg} \\
 1 \text{ J} = 6.242 \times 10^{18} \text{ eV} & 1 \text{ eV} = 1.602 \times 10^{19} \text{ J} \\
 1 \text{ m} = 1 \times 10^{10} \text{ Å} & 1 \text{ Å} = 1 \times 10^{-10} \text{ m}
 \end{array}$$

Periodic Table of the Elements

1 H	$^1S_{1/2}$ Hydrogen 1.008 $1s$	2 He	1S_0 Helium 4.0026 $1s^2$ 24.8374
3 Li	$^1S_{1/2}$ Lithium 0.94 $1s^2 2s^1$ 3.3227	4 Be	1S_0 Boron 9.0122 $1s^2 2s^2$ 9.3917
11 Na	$^1S_{1/2}$ Sodium 22.90 (He)3s ¹ 5.1391	12 Mg	1S_0 Magnesium 24.305 (He)3s ² 7.6462
19 K	$^1S_{1/2}$ Potassium 39.098 (He)3s ¹ 4.3407	20 Ca	1S_0 Calcium 40.078 (He)3s ² 6.1132
37 Rb	$^1S_{1/2}$ Rubidium 85.468 (He)3s ² 4.1771	38 Sr	$^1S_{1/2}$ Strontium 87.62 (He)3s ² 5.6949
55 Cs	$^1S_{1/2}$ Cesium 132.91 (He)6s ¹ 3.8939	56 Ba	1S_0 Barium 133.3 (He)6s ² 5.2117
87 Fr	$^1S_{1/2}$ Francium (He)7s ¹ 4.0727	88 Ra	1S_0 Radium (He)7s ² 5.2784
5 Li	$^3S_{1/2}$ Lithium 0.94 $1s^2 2s^1$ 3.3227	6 Be	3P_0 Boron 9.0122 $1s^2 2s^2 p^0$ 9.3917
11 Na	$^3S_{1/2}$ Sodium 22.90 (He)3s ¹ 5.1391	12 Mg	$^3S_{1/2}$ Magnesium 24.305 (He)3s ² 7.6462
19 K	$^3S_{1/2}$ Potassium 39.098 (He)3s ¹ 4.3407	20 Ca	$^3S_{1/2}$ Calcium 40.078 (He)3s ² 6.1132
37 Rb	$^3S_{1/2}$ Rubidium 85.468 (He)3s ² 4.1771	38 Sr	$^3S_{1/2}$ Strontium 87.62 (He)3s ² 5.6949
55 Cs	$^3S_{1/2}$ Cesium 132.91 (He)6s ¹ 3.8939	56 Ba	$^3S_{1/2}$ Barium 133.3 (He)6s ² 5.2117
87 Fr	$^3S_{1/2}$ Francium (He)7s ¹ 4.0727	88 Ra	$^3S_{1/2}$ Radium (He)7s ² 5.2784
10 B	3P_0 Boron 9.0122 $1s^2 2s^2 p^0$ 9.3917	11 C	3P_0 Carbon 12.011 $1s^2 2s^2 p^0$ 10.81
13 Al	3P_1 Aluminum 26.982 (He)3s ² p ¹ 8.1517	14 Si	3P_0 Silicon 28.085 (He)3s ² p ² 8.1517
16 S	3P_1 Sulfur 32.085 (He)3s ² p ⁴ 10.8300	17 Cl	3P_1 Chlorine 30.974 (He)3s ² p ⁵ 10.8687
18 Ar	3P_1 Argon 39.948 (Ne)3s ² 3p ⁴ 15.7596	36 Ge	3P_2 Germanium 72.692 (Ar)3d ¹⁰ 4s ² 5.9993
34 Se	3P_2 Selenium 78.924 (Ar)3d ¹⁰ 4s ² 4p ⁴ 11.8138	33 As	3P_2 Arsenic 74.922 (Ar)3d ¹⁰ 4s ² 4p ³ 7.8994
36 Kr	3P_2 Krypton 83.798 (Ar)3d ¹⁰ 4s ² 4p ⁶ 13.9996	31 Ga	3P_2 Gallium 69.723 (Ar)3d ¹⁰ 4s ² 4p ³ 7.7264
50 Sn	3P_2 Tin 127.60 (K)3d ¹⁰ 5s ² 5p ⁴ 9.0997	49 Cd	3P_2 Cadmium 121.70 (K)3d ¹⁰ 5s ² 5p ³ 7.3439
52 Te	3P_2 Tellurium 127.60 (K)3d ¹⁰ 5s ² 5p ⁴ 10.4513	51 Sb	3P_2 Antimony 131.29 (K)3d ¹⁰ 5s ² 5p ⁵ 12.1298
54 Xe	3P_2 Xenon 131.29 (K)3d ¹⁰ 5s ² 5p ⁶ 12.1298	47 Ag	3P_2 Silver 107.87 (K)3d ¹⁰ 5s ² 8.3369
56 Po	3P_2 Poison 208.98 (Hg)6s ² 8.4114	46 Sc	3P_2 Scandium 101.07 (K)3d ¹⁰ 5s ² 7.3405
58 Ru	3P_2 Rhodium 102.91 (K)3d ¹⁰ 5s ² 7.4589	45 Rh	3P_2 Rhodium 102.91 (K)3d ¹⁰ 5s ² 7.4589
59 Os	3P_2 Osmium 190.23 (He)4f ¹³ 5s ² 8.4382	43 Tc	3P_2 Technetium 101.07 (K)3d ¹⁰ 5s ² 7.3405
60 Pt	3P_2 Platinum 192.22 (K)3d ¹⁰ 5s ² 8.9558	41 Nb	3P_2 Niobium 92.906 (K)3d ¹⁰ 5s ² 7.1194
61 Cu	3P_2 Copper 63.58 (Ar)3d ¹⁰ 4s ² 7.6850	28 Fe	3P_2 Iron 56.745 (Ar)3d ⁶ 4s ² 7.9025
62 Ag	3P_2 Silver 106.42 (K)3d ¹⁰ 5s ² 7.5762	27 Co	3P_2 Cobalt 55.933 (Ar)3d ⁷ 4s ² 7.8810
63 Au	3P_2 Gold 196.97 (K)3d ¹⁰ 5s ² 8.9558	26 Fe	3P_2 Iron 55.938 (Ar)3d ⁶ 4s ² 7.9025
64 Gd	3P_2 Gadolinium 151.96 (Ce)4f ⁷ 5d ¹ 6s ² 5.6437	25 Ti	3P_2 Titanium 51.996 (Ar)3d ² 4s ² 7.6765
65 Tb	3P_2 Terbium 157.25 (Ce)4f ¹³ 5d ² 6.1498	24 Cr	3P_2 Chromium 54.938 (Ar)3d ⁵ 4s ² 7.8430
66 Dy	3P_2 Dysprosium 162.50 (Ce)4f ¹² 5d ¹ 5.5391	23 V	3P_2 Vanadium 50.945 (Ar)3d ³ 4s ² 7.8430
67 Ho	3P_2 Holmium 164.93 (Ce)4f ¹² 5d ¹ 6.0205	22 Ni	3P_2 Nickel 58.63 (Ar)3d ⁸ 4s ² 7.6850
68 Er	3P_2 Erbium 167.26 (Ce)4f ¹² 5d ² 6.1843	21 Sc	3P_2 Scandium 55.938 (Ar)3d ¹ 4s ² 7.7665
69 Tm	3P_2 Thulium 174.97 (Ce)4f ¹² 5d ³ 6.4269	20 Ca	3P_2 Calcium 55.938 (Ar)3d ² 4s ² 7.8430
70 Lu	3P_2 Lutetium 175.05 (Ce)4f ¹² 5d ⁴ 6.8143	19 K	3P_2 Potassium 55.938 (Ar)3d ¹ 4s ² 7.8430
71 Hg	3P_2 Mercury 200.59 (Hg)6s ² 10.4375	18 Ar	3P_2 Argon 107.22 (Hg)6s ² 10.4513
72 W	3P_2 Tungsten 192.22 (K)3d ¹⁰ 5s ² 8.9558	17 F	3P_2 Fluorine (C)2s ¹ 9.8030
73 Os	3P_2 Osmium 190.23 (K)3d ¹⁰ 5s ² 8.8335	16 S	3P_2 Sulfur 32.085 (He)3s ² 8.1517
74 Ir	3P_2 Iridium 198.21 (K)3d ¹⁰ 5s ² 8.7840	15 P	3P_2 Phosphorus 30.974 (He)3s ² 8.1067
75 Re	3P_2 Rhenium 183.84 (K)3d ¹⁰ 5s ² 8.7840	14 F	3P_2 Fluorine (C)2s ¹ 9.8030
76 Pt	3P_2 Platinum 192.22 (K)3d ¹⁰ 5s ² 8.9558	13 Ne	3P_2 Nitrogen 14.007 (He)3s ² 8.0727
77 Au	3P_2 Gold 196.97 (K)3d ¹⁰ 5s ² 8.9558	12 He	3P_2 Helium 4.0026 $1s^2$ 24.8374
78 Hg	3P_2 Mercury 200.59 (Hg)6s ² 10.4375	11 O	3P_2 Oxygen 16.998 $1s^2 2s^2$ 17.4228
79 Os	3P_2 Osmium 151.96 (Ce)4f ¹³ 5d ¹ 5.6437	10 Ne	3P_2 Neon 20.180 $1s^2 2s^2 2p^6$ 21.6405
80 Pt	3P_2 Platinum 192.22 (K)3d ¹⁰ 5s ² 8.9558	9 F	3P_2 Fluorine 18.998 $1s^2 2s^2 2p^5$ 17.4228
81 Au	3P_2 Gold 196.97 (K)3d ¹⁰ 5s ² 8.9558	8 O	3P_2 Oxygen 16.998 $1s^2 2s^2 2p^4$ 13.8181
82 Pb	3P_2 Lead 208.98 (Hg)6s ² 10.4375	7 N	3P_2 Nitrogen 14.007 $1s^2 2s^2 2p^3$ 12.011
83 Bi	3P_2 Bismuth 214.97 (Hg)6s ² 10.4375	6 C	3P_0 Carbon 12.011 $1s^2 2s^2 p^0$ 10.81
84 Te	3P_2 Tellurium 127.60 (K)3d ¹⁰ 5s ² 9.0997	5 B	3P_0 Boron 10.81 $1s^2 2s^2 p^0$ 8.2990
85 Bi	3P_2 Bismuth 208.98 (Hg)6s ² 8.4114	4 Be	3P_0 Boron 9.0122 $1s^2 2s^2$ 9.3917
86 Rn	3P_2 Radarion (He)222 10.4513	3 He	3P_0 Helium 4.0026 $1s^2$ 24.8374
87 Fr	3P_2 Francium (He)7s ¹ 4.0727	2 He	3P_0 Helium 4.0026 $1s^2$ 24.8374
88 Ra	3S_0 Radium (He)7s ² 5.2784	1 H	$^1S_{1/2}$ Hydrogen 1.008 $1s$

Figure A.1: Periodic table of the elements: ● Metals; ● Metaloids; ● Nonmetals.

B

Integrals

B.1 $\int_0^\infty x^{p-1}/(e^x - 1) dx$

integral from 0 to infinity of x to the p-1 over e to the x - 1 dx Consider the integral

$$I_p = \int_0^\infty \frac{x^3}{e^x - 1} dx \quad (\text{B.1.1})$$

for p with $\text{Re}(p) > 1$. Rewriting the integrand we can identify a geometric series:

$$I_p = \int_0^\infty e^{-x} x^{p-1} \frac{1}{1 - e^{-x}} dx \quad (\text{B.1.2})$$

$$= \int_0^\infty e^{-x} x^{p-1} \left[\sum_{m=0}^{\infty} (e^{-x})^m \right] dx \quad (\text{B.1.3})$$

$$= \int_0^\infty x^{p-1} \sum_{m=0}^{\infty} e^{-(m+1)x}. \quad (\text{B.1.4})$$

The geometric series converges uniformly in its region of convergence, which is everywhere in this integral since $x > 0$ and so $|e^{-x}| < 1$. This means we can swap the sum and integral to get

$$I_p = \sum_{m=0}^{\infty} \int_0^\infty e^{-(m+1)x} x^{p-1} dx. \quad (\text{B.1.5})$$

We can re-index the sum to $n = m + 1$ to give

$$I_p = \sum_{n=1}^{\infty} \int_0^\infty e^{-nx} x^{p-1} dx. \quad (\text{B.1.6})$$

We now make the substitution $u = nx$, so $du = n dx$. This gives

$$I_p = \sum_{n=1}^{\infty} \frac{1}{n} \int_0^\infty e^{-u} \frac{u^{p-1}}{n^{p-1}} du \quad (\text{B.1.7})$$

$$= \sum_{n=1}^{\infty} \frac{1}{n^p} \int_0^\infty e^{-u} u^{p-1} du \quad (\text{B.1.8})$$

$$= \zeta(p) \Gamma(p) \quad (\text{B.1.9})$$

where we have identified the **Riemann zeta function** and **gamma function**

$$\zeta(p) := \sum_{n=1}^{\infty} \frac{1}{n^p}, \quad \text{and} \quad \Gamma(p) := \int_0^{\infty} e^{-u} u^{p-1} du. \quad (\text{B.1.10})$$

Both of these can be evaluated exactly only at certain values. One of the values we care about is $p = 4$, for which we have $\zeta(4) = \pi^4/90$ and $\Gamma(4) = (4-1)! = 6$. This gives

$$I_4 = \int_0^{\infty} \frac{x^3}{e^x - 1} dx = \zeta(4)\Gamma(4) = \frac{\pi^4}{15}. \quad (\text{B.1.11})$$

Bibliography

- [1] Inductiveload, *Brillouin Zone (1st, FCC)*, [https://commons.wikimedia.org/wiki/File:Brillouin_Zone_\(1st,_FCC\).svg](https://commons.wikimedia.org/wiki/File:Brillouin_Zone_(1st,_FCC).svg), 05/05/2008 (accessed 01/11/2021).
- [2] www.phonon.fc.pl (no longer available) via course notes (accessed 01/11/2021).

Acronyms

B

BCC: body-centred cubic [10](#)

C

CCP: cubic close-packed [11](#)

CGS: centimetre, gram, second [30](#)

D

DFT: density functional theory [46](#)

F

FCC: face-centred cubic [10](#)

H

HCP: hexagonal close-packed [12](#)

L

LED: light emitting diode [39](#)

S

SANS: small angle neutron scattering [26](#)

SAXS: small angle x-ray scattering [26](#)

SI: système international [30](#)

Z

ZPE: zero-point energy [29](#)

Index

A

absorbance, 101
absorption coefficient, 101
adiabatic approximation, *see* Born–Oppenheimer approximation
atomic form factor, 22

B

Bloch’s theorem, 88
Born–Oppenheimer approximation, 65
Born–von Karman boundary condition, 67
Bragg’s law, 16, 21
Bravais lattices, 8
Brillouin zone, 45

C

central equation, 92
centre of inversion, 7
CGS, 30
chemical potential, 69
coefficient of linear thermal expansion, 61
cohesive energy, 28
coordination number, 35
crystal, 3
cubic close packed, 11
cyclotron frequency, 80

D

Debye
frequency, 54
temperature, 55
wave vector, 54
Debye model, 53
density functional theory, 46
diffraction, 16
direct band gap, 101

dispersion interaction, *see* van der Waals
doping, 110
drift velocity, 76
Drude model, 76
Drude–Sommerfeld model, 107
Dulong–Petit law, 42

E

effective mass, 105
tensor, 106
Einstein model, 42
Einstein temperature, 43
empty lattice approximation, 85
energy band, 82

F

Fermi
energy, 68
level, 70
sphere, 68
surface, 92
temperature, 68
velocity, 68
wave vector, 68
Fermi–Dirac distribution, 69
free electron approximation, 65

G

gamma function, 117
geometrical scattering, 62
group velocity, 49

H

Hall
coefficeint, 80
effect, 79
voltage, 79
hole, 103

I

- independent electron approximation, 65
 indirect band gap, 101
 induced dipole-dipole interaction, *see* van der Waals
 intrinsic semiconductor, 106
 inversion symmetry, 7

K

- Kronig–Penny model, 85

L

- lattice
 point, 3
 vector, 3
 lattice sums, 33
 lattice, crystal, 3
 Laue diffraction condition, 21
 law of mass action, 110
 Lennard-Jones potential, 32
 London interaction, *see* van der Waals
 Lorenz number, 79

M

- Madelung constant, 37
 Madelung energy, 38
 Miller indices, 13
 mirror plane symmetry, 5
 mobility, 107

N

- normal scattering, 62

P

- periodic crystal, 3
 phonon dispersion relation, 45
 primitive unit cell, 4

Q

- quasicrystals, 26

R

- reciprocal lattice, 18
 reciprocal lattice vectors, 19
 Riemann zeta function, 117
 rotational symmetry, 6

S

- scattering amplitude, 17
 scattering vector, 20

semiconductor, 100

semimetal, 100
 Sommerfeld constant, 71
 speed of sound, 49
 structure factor, 22
Strukturbericht symbols, 39
 systematic absence, 24

T

thermal effective mass, 75
 transmittance, 101

U

umklapp scattering, 62
 unit cell, 4

V

van der Waals, 29

W

Widemann–Franz law, 79
 Wigner–Seitz construction, 45