

Lecture Notes for Solid State Physics
(3rd Year Course 6)
Hilary Term 2012

©Professor Steven H. Simon
Oxford University

January 9, 2012

Short Preface to My Second Year Lecturing This Course

Last year was my first year teaching this course. In fact, it was my first experience teaching any undergraduate course. I admit that I learned quite a bit from the experience. The good news is that the course was viewed mostly as a success, even by the tough measure of student reviews. I particularly would like to thank that student who wrote on his or her review that I deserve a raise — and I would like to encourage my department chair to post this review on his wall and refer to it frequently.

With luck, the second iteration of the course will be even better than the first. Having learned so much from teaching the course last year, I hope to improve it even further for this year. One of the most important things I learned was how much students appreciate a clear, complete, and error-free set of notes. As such, I am spending quite a bit of time reworking these notes to make them as perfect as possible.

Repeating my plea from last year, if you can think of ways that these notes (or this course) could be further improved (correction of errors or whatnot) please let me know. The next generation of students will certainly appreciate it and that will improve your Karma. ☺

Oxford, United Kingdom
January, 2012.

Preface

When I was an undergraduate, I thought solid state physics (a sub-genre of condensed matter physics) was perhaps the worst subject that any undergraduate could be forced to learn – boring and tedious, “squalid state” as it was commonly called¹. How much would I really learn about the universe by studying the properties of crystals? I managed to avoid taking this course altogether. My opinion at the time was not a reflection of the subject matter, but rather was a reflection of how solid state physics was taught.

Given my opinion as an undergraduate, it is a bit ironic that I have become a condensed matter physicist. But once I was introduced to the subject properly, I found that condensed matter was my favorite subject in all of physics – full of variety, excitement, and deep ideas. Many many physicists have come to this same conclusion. In fact, condensed matter physics is by far the largest single subfield of physics (the annual meeting of condensed matter physicists in the United States attracts over 6000 physicists each year!). Sadly a first introduction to the topic can barely scratch the surface of what constitutes the broad field of condensed matter.

Last year when I was told that a new course was being prepared to teach condensed matter physics to third year Oxford undergraduates, I jumped at the opportunity to teach it. I felt that it *must* be possible to teach a condensed matter physics course that is just as interesting and exciting as any other course that an undergraduate will ever take. It must be possible to convey the excitement of real condensed matter physics to the undergraduate audience. I hope I will succeed in this task. You can judge for yourself.

The topics I was asked to cover (being given little leeway in choosing the syllabus) are not atypical for a solid state physics course. In fact, the new condensed matter syllabus is extremely similar to the old Oxford B2 syllabus – the main changes being the removal of photonics and device physics. A few other small topics, such as superconductivity and point-group symmetries, are also nonexaminable now, or are removed altogether. A few other topics (thermal expansion, chemical bonding) are now added by mandate of the IOP².

At any rate, the changes to the old B2 syllabus are generally minor, so I recommend that Oxford students use the old B2 exams as a starting point for figuring out what it is they need to study as the exams approach. In fact, I have used precisely these old exams to figure out what I need to teach. Being that the same group of people will be setting the exams this year as set them last year, this seems like a good idea. As with most exams at Oxford, one starts to see patterns in terms of what type of questions are asked year after year. The lecture notes contained here are designed to cover exactly this crucial material. I realize that these notes are a lot of material, and for this I apologize. However, this is the *minimum* set of notes that covers all of the topics that have shown up on old B2 exams. The actual lectures for this course will try to cover everything in these notes, but a few of the less crucial pieces will necessarily be glossed over in the interest of time.

Many of these topics are covered well in standard solid state physics references that one might find online, or in other books. The reason I am giving these lectures (and not just telling students to go read a standard book) is because condensed matter/solid-state is an enormous subject — worth many years of lectures — and one needs a guide to decide what subset of topics

¹This jibe against solid state physics can be traced back to the Nobel Laureate Murray Gell-Mann, discoverer of the quark, who famously believed that there was nothing interesting in any endeavor but particle physics. Interestingly he now studies complexity — a field that mostly arose from condensed matter.

²We can discuss elsewhere whether or not we should pay attention to such mandates in general – although these particular mandates do not seem so odious.

are most important (at least in the eyes of the examination committee). I believe that the lectures contained here give depth in some topics, and gloss over other topics, so as to reflect the particular topics that are deemed important at Oxford. These topics may differ a great deal from what is deemed important elsewhere. In particular, Oxford is extremely heavy on scattering theory (x-ray and neutron diffraction) compared with most solid state courses or books that I have seen. But on the other hand, Oxford does not appear to believe in group representations (which resulted in my elimination of point group symmetries from the syllabus).

I cannot emphasize enough that there are many many extremely good books on solid-state and condensed matter physics already in existence. There are also many good resources online (including the rather infamous “Britney Spears’ guide to semiconductor physics” — which is tongue-in-cheek about Britney Spears³, but actually is a very good reference about semiconductors). I will list here some of the books that I think are excellent, and throughout these lecture notes, I will try to point you to references that I think are helpful.

- **States of Matter**, by David L. Goodstein, Dover
Chapter 3 of this book is a very brief but well written and easy to read description of much of what we will need to cover (but not all, certainly). The book is also published by Dover which means it is super-cheap in paperback. Warning: It uses cgs units rather than SI units, which is a bit annoying.
- **Solid State Physics, 2nd ed** by J. R. Hook and H. E. Hall, Wiley
This is frequently the book that students like the most. It is a first introduction to the subject and is much more introductory than Ashcroft and Mermin.
- **The Solid State**, by H M Rosenberg, OUP
This slightly more advanced book was written a few decades ago to cover what was the solid state course at Oxford at that time. Some parts of the course have since changed, but other parts are well covered in this book.
- **Solid-State Physics, 4ed**, by H. Ibach and H. Luth, Springer-Verlag
Another very popular book on the subject, with quite a bit of information in it. More advanced than Hook and Hall
- **Solid State Physics**, by N. W. Ashcroft and D. N. Mermin, Holt-Sanders
This is the standard complete introduction to solid state physics. It has many many chapters on topics we won’t be studying, and goes into great depth on almost everything. It may be a bit overwhelming to try to use this as a reference because of information-overload, but it has good explanations of almost everything. On the whole, this is my favorite reference. Warning: Also uses cgs units.
- **Introduction to Solid State Physics, 8ed**, by Charles Kittel⁴, Wiley
This is a classic text. It gets mixed reviews by some as being unclear on many matters. It is somewhat more complete than Hooke and Hall, less so than Ashcroft and Mermin. Its selection of topics and organization may seem a bit strange in the modern era.
- **The Basics of Crystallography and Diffraction, 3ed**, by C Hammond, OUP
This book has historically been part of the syllabus, particularly for the scattering theory part of the course. I don’t like it much.

³This guide was written when Ms. Spears was just a popular young performer and not the complete train wreck that she appears to be now.

⁴Kittel happens to be my dissertation-supervisor’s dissertation-supervisor’s dissertation-supervisor’s dissertation-supervisor, for whatever that is worth.

- **Structure and Dynamics**, by M.T. Dove, Oxford University Press
This is a more advanced book that covers scattering in particular. It is used in the Condensed Matter option 4-th year course.
- **Magnetism in Condensed Matter**, by Stephen Blundell, OUP
Well written advanced material on the magnetism part of the course. It is used in the Condensed Matter option 4-th year course.
- **Band Theory and Electronic Properties of Solids**, by John Singleton, OUP
More advanced material on electrons in solids. Also used in the Condensed Matter option 4-th year course.
- **Solid State Physics**, by G. Burns, Academic
Another more advanced book. Some of its descriptions are short but very good.

I will remind my reader that these notes are a first draft. I apologize that they do not cover the material uniformly. In some places I have given more detail than in others – depending mainly on my enthusiasm-level at the particular time of writing. I hope to go back and improve the quality as much as possible. Updated drafts will hopefully be appearing.

Perhaps this pile of notes will end up as a book, perhaps they will not. This is not my point. My point is to write something that will be helpful for this course. If you can think of ways that these notes could be improved (correction of errors or whatnot) please let me know. The next generation of students will certainly appreciate it and that will improve your Karma. ☺

Oxford, United Kingdom
January, 2011.

Acknowledgements

Needless to say, I pilfered a fair fraction of the content of this course from parts of other books (mostly mentioned above). The authors of these books put great thought and effort into their writing. I am deeply indebted to these giants who have come before me. Additionally, I have stolen many ideas about how this course should be taught from the people who have taught the course (and similar courses) at Oxford in years past. Most recently this includes Mike Glazer, Andrew Boothroyd, and Robin Nicholas.

I am also very thankful for all the people who have helped me proofread, correct, and otherwise tweak these notes and the homework problems. These include in particular Mike Glazer, Alex Hearmon, Simon Davenport, Till Hackler, Paul Stubbley, Stephanie Simmons, Katherine Dunn, and Joost Slingerland.

Finally, I thank my father for helping proofread and improve these notes... and for a million other things.

Contents

1	About Condensed Matter Physics	1
1.1	What is Condensed Matter Physics	1
1.2	Why Do We Study Condensed Matter Physics?	1
I	Physics of Solids without Considering Microscopic Structure: The Early Days of Solid State	5
2	Specific Heat of Solids: Boltzmann, Einstein, and Debye	7
2.1	Einstein's Calculation	8
2.2	Debye's Calculation	11
2.2.1	About Periodic (Born-Von-Karman) Boundary Conditions	12
2.2.2	Debye's Calculation Following Planck	13
2.2.3	Debye's "Interpolation"	15
2.2.4	Some Shortcomings of the Debye Theory	15
2.3	Summary of Specific Heat of Solids	17
2.4	Appendix to this Chapter: $\zeta(4)$	17
3	Electrons in Metals: Drude Theory	19
3.1	Electrons in Fields	20
3.1.1	Electrons in an Electric Field	20
3.1.2	Electrons in Electric and Magnetic Fields	21
3.2	Thermal Transport	23
3.3	Summary of Drude Theory	25
4	More Electrons in Metals: Sommerfeld (Free Electron) Theory	27
4.1	Basic Fermi-Dirac Statistics	28

4.2	Electronic Heat Capacity	30
4.3	Magnetic Spin Susceptibility (Pauli Paramagnetism)	32
4.4	Why Drude Theory Works so Well	35
4.5	Shortcomings of the Free Electron Model	35
4.6	Summary of (Sommerfeld) Free Electron Theory	37
II Putting Materials Together		39
5	What Holds Solids Together: Chemical Bonding	41
5.1	General Considerations about Bonding	41
5.2	Ionic Bonds	44
5.3	Covalent Bond	47
5.3.1	Particle in a Box Picture	47
5.3.2	Molecular Orbital or Tight Binding Theory	47
5.4	Van der Waals, Fluctuating Dipole Forces, or Molecular Bonding	53
5.5	Metallic Bonding	54
5.6	Hydrogen bonds	55
5.7	Summary of Bonding (Pictorial)	55
6	Types of Matter	57
III Toy Models of Solids in One Dimension		61
7	One Dimensional Model of Compressibility, Sound, and Thermal Expansion	63
8	Vibrations of a One Dimensional Monatomic Chain	67
8.1	First Exposure to the Reciprocal Lattice	68
8.2	Properties of the Dispersion of the One Dimensional Chain	70
8.3	Quantum Modes: Phonons	72
8.4	Crystal Momentum	74
8.5	Summary of Vibrations of the One Dimensional Monatomic Chain	75
9	Vibrations of a One Dimensional Diatomic Chain	77
9.1	Diatomic Crystal Structure: Some useful definitions	77
9.2	Normal Modes of the Diatomic Solid	79

9.3	Summary of Vibrations of the One Dimensional Diatomic Chain	85
10	Tight Binding Chain (Interlude and Preview)	87
10.1	Tight Binding Model in One Dimension	87
10.2	Solution of the Tight Binding Chain	89
10.3	Introduction to Electrons Filling Bands	92
10.4	Multiple Bands	93
10.5	Summary of Tight Binding Chain	95
IV	Geometry of Solids	97
11	Crystal Structure	99
11.1	Lattices and Unit Cells	99
11.2	Lattices in Three Dimensions	106
11.3	Summary of Crystal Structure	112
12	Reciprocal Lattice, Brillouin Zone, Waves in Crystals	115
12.1	The Reciprocal Lattice in Three Dimensions	115
12.1.1	Review of One Dimension	115
12.1.2	Reciprocal Lattice Definition	116
12.1.3	The Reciprocal Lattice as a Fourier Transform	117
12.1.4	Reciprocal Lattice Points as Families of Lattice Planes	118
12.1.5	Lattice Planes and Miller Indices	120
12.2	Brillouin Zones	123
12.2.1	Review of One Dimensional Dispersions and Brillouin Zones	123
12.2.2	General Brillouin Zone Construction	124
12.3	Electronic and Vibrational Waves in Crystals in Three Dimensions	125
12.4	Summary of Reciprocal Space and Brillouin Zones	127
V	Neutron and X-Ray Diffraction	129
13	Wave Scattering by Crystals	131
13.1	The Laue and Bragg Conditions	132
13.1.1	Fermi's Golden Rule Approach	132
13.1.2	Diffraction Approach	133

13.1.3	Equivalence of Laue and Bragg conditions	134
13.2	Scattering Amplitudes	135
13.2.1	Systematic Absences and More Examples	138
13.3	Methods of Scattering Experiments	140
13.3.1	Advanced Methods (interesting and useful but you probably won't be tested on this)	140
13.3.2	Powder Diffraction (you will almost certainly be tested on this!)	141
13.4	Still more about scattering	147
13.4.1	Variant: Scattering in Liquids and Amorphous Solids	147
13.4.2	Variant: Inelastic Scattering	148
13.4.3	Experimental Apparatus	148
13.5	Summary of Diffraction	149
VI	Electrons in Solids	151
14	Electrons in a Periodic Potential	153
14.1	Nearly Free Electron Model	153
14.1.1	Degenerate Perturbation Theory	155
14.2	Bloch's Theorem	160
14.3	Summary of Electrons in a Periodic Potential	161
15	Insulator, Semiconductor, or Metal	163
15.1	Energy Bands in One Dimension: Mostly Review	163
15.2	Energy Bands in Two (or More) Dimensions	166
15.3	Tight Binding	168
15.4	Failures of the Band-Structure Picture of Metals and Insulators	170
15.5	Band Structure and Optical Properties	171
15.5.1	Optical Properties of Insulators and Semiconductors	171
15.5.2	Direct and Indirect Transitions	171
15.5.3	Optical Properties of Metals	172
15.5.4	Optical Effects of Impurities	173
15.6	Summary of Insulators, Semiconductors, and Metals	174
16	Semiconductor Physics	175
16.1	Electrons and Holes	175

16.1.1	Drude Transport: Redux	178
16.2	Adding Electrons or Holes With Impurities: Doping	179
16.2.1	Impurity States	180
16.3	Statistical Mechanics of Semiconductors	183
16.4	Summary of Statistical Mechanics of Semiconductors	187
17	Semiconductor Devices	189
17.1	Band Structure Engineering	189
17.1.1	Designing Band Gaps	189
17.1.2	Non-Homogeneous Band Gaps	190
17.1.3	Summary of the Examinable Material	190
17.2	p - n Junction	191
VII	Magnetism and Mean Field Theories	193
18	Magnetic Properties of Atoms: Para- and Dia-Magnetism	195
18.1	Basic Definitions of types of Magnetism	196
18.2	Atomic Physics: Hund's Rules	197
18.2.1	Why Moments Align	200
18.3	Coupling of Electrons in Atoms to an External Field	202
18.4	Free Spin (Curie or Langevin) Paramagnetism	204
18.5	Larmor Diamagnetism	206
18.6	Atoms in Solids	207
18.6.1	Pauli Paramagnetism in Metals	207
18.6.2	Diamagnetism in Solids	207
18.6.3	Curie Paramagnetism in Solids	208
18.7	Summary of Atomic Magnetism; Paramagnetism and Diamagnetism	209
19	Spontaneous Order: Antiferro-, Ferri-, and Ferro-Magnetism	211
19.1	(Spontaneous) Magnetic Order	212
19.1.1	Ferromagnets	212
19.1.2	Antiferromagnets	212
19.1.3	Ferrimagnetism	214
19.2	Breaking Symmetry	214
19.2.1	Ising Model	215

19.3 Summary of Magnetic Orders	216
20 Domains and Hysteresis	217
20.1 Macroscopic Effects in Ferromagnets: Domains	217
20.1.1 Disorder and Domain Walls	218
20.1.2 Disorder Pinning	219
20.1.3 The Bloch/Néel Wall	219
20.2 Hysteresis in Ferromagnets	222
20.2.1 Single-Domain Crystallites	222
20.2.2 Domain Pinning and Hysteresis	223
20.3 Summary of Domains and Hysteresis in Ferromagnets	224
21 Mean Field Theory	227
21.1 Mean Field Equations for the Ferromagnetic Ising Model	227
21.2 Solution of Self-Consistency Equation	229
21.2.1 Paramagnetic Susceptibility	231
21.2.2 Further Thoughts	232
21.3 Summary of Mean Field Theory	232
22 Magnetism from Interactions: The Hubbard Model	235
22.1 Ferromagnetism in the Hubbard Model	236
22.1.1 Hubbard Ferromagnetism Mean Field Theory	236
22.1.2 Stoner Criterion	237
22.2 Mott Antiferromagnetism in the Hubbard Model	239
22.3 Summary of the Hubbard Model	241
22.4 Appendix: The Hubbard model for the Hydrogen Molecule	241
23 Magnetic Devices	245
Indices	247
Index of People	248
Index of Topics	250

Chapter 1

About Condensed Matter Physics

This chapter is just my personal take on why this topic is interesting. It seems unlikely to me that any exam would ask you why you study this topic, so you should probably consider this section to be not examinable. Nonetheless, you might want to read it to figure out why you should think this course is interesting if that isn't otherwise obvious.

1.1 What is Condensed Matter Physics

Quoting Wikipedia:

Condensed matter physics is the field of physics that deals with the macroscopic and microscopic physical properties of matter. In particular, it is concerned with the “condensed” phases that appear whenever the number of constituents in a system is extremely large and the interactions between the constituents are strong. The most familiar examples of condensed phases are solids and liquids, which arise from the electromagnetic forces between atoms.

The use of the term “Condensed Matter” being more general than just solid state was coined and promoted by Nobel-Laureate Philip W. Anderson.

1.2 Why Do We Study Condensed Matter Physics?

There are several very good answers to this question

1. Because it is the world around us

Almost all of the physical world that we see is in fact condensed matter. Questions such as

- why are metals shiny and why do they feel cold?
- why is glass transparent?

- why is water a fluid, and why does fluid feel wet?
- why is rubber soft and stretchy?

These questions are all in the domain of condensed matter physics. In fact almost every question you might ask about the world around you, short of asking about the sun or stars, is probably related to condensed matter physics in some way.

2. Because it is useful

Over the last century our command of condensed matter physics has enabled us humans to do remarkable things. We have used our knowledge of physics to engineer new materials and exploit their properties to change our world and our society completely. Perhaps the most remarkable example is how our understanding of solid state physics enabled new inventions exploiting semiconductor technology, which enabled the electronics industry, which enabled computers, iPhones, and everything else we now take for granted.

3. Because it is deep

The questions that arise in condensed matter physics are as deep as those you might find anywhere. In fact, many of the ideas that are now used in other fields of physics can trace their origins to condensed matter physics.

A few examples for fun:

- The famous Higgs boson, which the LHC is searching for, is no different from a phenomenon that occurs in superconductors (the domain of condensed matter physicists). The Higgs mechanism, which gives mass to elementary particles is frequently called the “Anderson-Higgs” mechanism, after the condensed matter physicist Phil Anderson (the same guy who coined the term “condensed matter”) who described much of the same physics before Peter Higgs, the high energy theorist.
- The ideas of the renormalization group (Nobel prize to Kenneth Wilson in 1982) was developed simultaneously in both high-energy and condensed matter physics.
- The ideas of topological quantum field theories, while invented by string theorists as theories of quantum gravity, have been discovered in the laboratory by condensed matter physicists!
- In the last few years there has been a mass exodus of string theorists applying black-hole physics (in N -dimensions!) to phase transitions in real materials. The very same structures exist in the lab that are (maybe!) somewhere out in the cosmos!

That this type of physics is deep is not just my opinion. The Nobel committee agrees with me. During this course we will discuss the work of no fewer than 50 Nobel laureates! (See the index of scientists at the end of this set of notes).

4. Because reductionism doesn’t work

begin{rant} People frequently have the feeling that if you continually ask “what is it made of” you learn more about something. This approach to knowledge is known as *reductionism*. For example, asking what water is made of, someone may tell you it is made from molecules, then molecules are made of atoms, atoms of electrons and protons, protons of quarks, and quarks are made of who-knows-what. But none of this information tells you anything about why water is wet, about why protons and neutrons bind to form nuclei, why the atoms bind to form water, and so forth. Understanding physics inevitably involves understanding how many objects all interact with each other. And this is where things get difficult very

quickly. We understand the Schroedinger equation extremely well for one particle, but the Schroedinger equations for four or more particles, while in principle solvable, in practice are never solved because they are too difficult — even for the world’s biggest computers. Physics involves figuring out what to do then. How are we to understand how many quarks form a nucleus, or how many electrons and protons form an atom if we cannot solve the many particle Schroedinger equation?

Even more interesting is the possibility that we understand very well the microscopic theory of a system, but then we discover that macroscopic properties *emerge* from the system that we did not expect. My personal favorite example is when one puts together many electrons (each with charge $-e$) one can sometimes find new particles emerging, each having one third the charge of an electron!¹ Reductionism would never uncover this — it misses the point completely. `end{rant}`

5. Because it is a Laboratory

Condensed matter physics is perhaps the best laboratory we have for studying quantum physics and statistical physics. Those of us who are fascinated by what quantum mechanics and statistical mechanics can do often end up studying condensed matter physics which is deeply grounded in both of these topics. Condensed matter is an infinitely varied playground for physicists to test strange quantum and statistical effects.

I view this entire course as an extension of what you have already learned in quantum and statistical physics. If you enjoyed those courses, you will likely enjoy this as well. If you did not do well in those courses, you might want to go back and study them again because many of the same ideas will arise here.

¹Yes, this truly happens. The Nobel prize in 1998 was awarded to Dan Tsui, Horst Stormer and Bob Laughlin, for discovery of this phenomenon known as the fractional quantum Hall effect.

Part I

Physics of Solids without Considering Microscopic Structure: The Early Days of Solid State

Chapter 2

Specific Heat of Solids: Boltzmann, Einstein, and Debye

Our story of condensed matter physics starts around the turn of the last century. It was well known (and you should remember from last year) that the heat capacity¹ of a monatomic (ideal) gas is $C_v = 3k_B/2$ per atom with k_B being Boltzmann's constant. The statistical theory of gases described why this is so.

As far back as 1819, however, it had also been known that for many solids the heat capacity is given by²

$$C = 3k_B \text{ per atom}$$

or

$$C = 3R$$

which is known as the *Law of Dulong-Petit*³. While this law is not always correct, it frequently is close to true. For example, at room temperature we have

With the exception of diamond, the law $C/R = 3$ seems to hold extremely well at room temperature, although at lower temperatures all materials start to deviate from this law, and typically

¹We will almost always be concerned with the heat capacity C per atom of a material. Multiplying by Avogadro's number gives the *molar heat capacity* or heat capacity per mole. The *specific heat* (denoted often as c rather than C) is the heat capacity per unit mass. However, the phrase "specific heat" is also used loosely to describe the molar heat capacity since they are both intensive quantities (as compared to the total heat capacity which is extensive — i.e., proportional to the amount of mass in the system). We will try to be precise with our language but one should be aware that frequently things are written in non-precise ways and you are left to figure out what is meant. For example, Really we should say C_v per atom = $3k_B/2$ rather than $C_v = 3k_B/2$ per atom, and similarly we should say C per mole = $3R$. To be more precise I really would have liked to title this chapter "Heat Capacity Per Atom of Solids" rather than "Specific Heat of Solids". However, for over a century people have talked about the "Einstein Theory of Specific Heat" and "Debye Theory of Specific Heat" and it would have been almost scandalous to not use this wording.

²Here I do not distinguish between C_p and C_v because they are very close to the same. Recall that $C_p - C_v = VT\alpha^2/\beta_T$ where β_T is the isothermal compressibility and α is the coefficient of thermal expansion. For a solid α is relatively small.

³Both Pierre Dulong and Alexis Petit were French chemists. Neither is remembered for much else besides this law.

Material	C/R
Aluminum	2.91
Antimony	3.03
Copper	2.94
Gold	3.05
Silver	2.99
Diamond	0.735

Table 2.1: Heat Capacities of Some Solids

C drops rapidly below some temperature. (And for diamond when the temperature is raised, the heat capacity increases towards $3R$ as well, see Fig. 2.2 below).

In 1896 Boltzmann constructed a model that accounted for this law fairly well. In his model, each atom in the solid is bound to neighboring atoms. Focusing on a single particular atom, we imagine that atom as being in a harmonic well formed by the interaction with its neighbors. In such a classical statistical mechanical model, the heat capacity of the vibration of the atom is $3k_B$ per atom, in agreement with Dulong-Petit. (Proving this is a good homework assignment that you should be able to answer with your knowledge of statistical mechanics and/or the equipartition theorem).

Several years later in 1907, Einstein started wondering about why this law does not hold at low temperatures (for diamond, “low” temperature appears to be room temperature!). What he realized is that quantum mechanics is important!

Einstein’s assumption was similar to that of Boltzmann. He assumed that every atom is in a harmonic well created by the interaction with its neighbors. Further he assumed that every atom is in an identical harmonic well and has an oscillation frequency ω (known as the “Einstein” frequency).

The quantum mechanical problem of a simple harmonic oscillator is one whose solution we know. We will now use that knowledge to determine the heat capacity of a single one dimensional harmonic oscillator. This entire calculation should look familiar from your statistical physics course.

2.1 Einstein’s Calculation

In one dimension, the eigenstates of a single harmonic oscillator are

$$E_n = \hbar\omega(n + 1/2)$$

with ω the frequency of the harmonic oscillator (the “Einstein frequency”). The partition function is then⁴

$$\begin{aligned} Z_{1D} &= \sum_{n \geq 0} e^{-\beta\hbar\omega(n+1/2)} \\ &= \frac{e^{-\beta\hbar\omega/2}}{1 - e^{-\beta\hbar\omega}} = \frac{1}{2 \sinh(\beta\hbar\omega/2)} \end{aligned}$$

⁴We will very frequently use the standard notation $\beta = 1/(k_B T)$.

The expectation of energy is then

$$\langle E \rangle = -\frac{1}{Z} \frac{\partial Z}{\partial \beta} = \frac{\hbar\omega}{2} \coth\left(\frac{\beta\hbar\omega}{2}\right) = \hbar\omega \left(n_B(\beta\hbar\omega) + \frac{1}{2} \right) \quad (2.1)$$

where n_B is the Bose⁵ occupation factor

$$n_B(x) = \frac{1}{e^x - 1}$$

This result is easy to interpret: the mode ω is an excitation that is excited on average n_B times, or equivalently there is a “boson” orbital which is “occupied” by n_B bosons.

Differentiating the expression for energy we obtain the heat capacity for a single oscillator,

$$C = \frac{\partial \langle E \rangle}{\partial T} = k_B (\beta\hbar\omega)^2 \frac{e^{\beta\hbar\omega}}{(e^{\beta\hbar\omega} - 1)^2}$$

Note that the high temperature limit of this expression gives $C = k_B$ (check this if it is not obvious!).

Generalizing to the three-dimensional case,

$$E_{n_x, n_y, n_z} = \hbar\omega[(n_x + 1/2) + (n_y + 1/2) + (n_z + 1/2)]$$

and

$$Z_{3D} = \sum_{n_x, n_y, n_z \geq 0} e^{-\beta E_{n_x, n_y, n_z}} = [Z_{1D}]^3$$

resulting in $\langle E_{3D} \rangle = 3\langle E_{1D} \rangle$, so correspondingly we obtain

$$C = 3k_B (\beta\hbar\omega)^2 \frac{e^{\beta\hbar\omega}}{(e^{\beta\hbar\omega} - 1)^2}$$

Plotted this looks like Fig. 2.1.

⁵Satyendra Bose worked out the idea of Bose statistics in 1924, but could not get it published until Einstein lent his support to the idea.

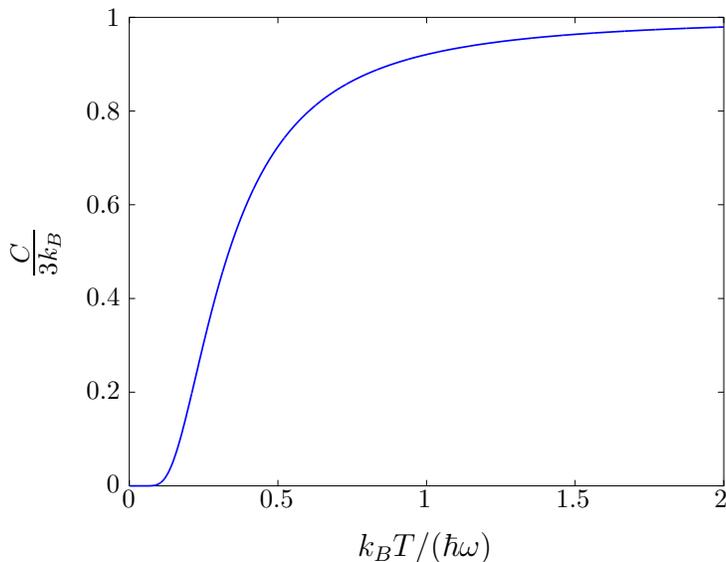


Figure 2.1: Einstein Heat Capacity Per Atom in Three Dimensions

Note that in the high temperature limit $k_B T \gg \hbar \omega$ recover the law of Dulong-Petit — $3k_B$ heat capacity per atom. However, at low temperature ($T \ll \hbar \omega / k_B$) the degrees of freedom “freeze out”, the system gets stuck in only the ground state eigenstate, and the heat capacity vanishes rapidly.

Einstein’s theory reasonably accurately explained the behavior of the the heat capacity as a function of temperature with only a single fitting parameter, the Einstein frequency ω . (Sometimes this frequency is quoted in terms of the Einstein temperature $\hbar \omega = k_B T_{Einstein}$). In Fig. 2.2 we show Einstein’s original comparison to the heat capacity of diamond.

For most materials, the Einstein frequency ω is low compared to room temperature, so the Dulong-Petit law hold fairly well (being relatively high temperature compared to the Einstein frequency). However, for diamond, ω is high compared to room temperature, so the heat capacity is lower than $3R$ at room temperature. The reason diamond has such a high Einstein frequency is that the bonding between atoms in diamond is very strong and its mass is relatively low (hence a high $\omega = \sqrt{\kappa/m}$ oscillation frequency with κ a spring constant and m the mass). These strong bonds also result in diamond being an exceptionally hard material.

Einstein’s result was remarkable, not only in that it explained the temperature dependence

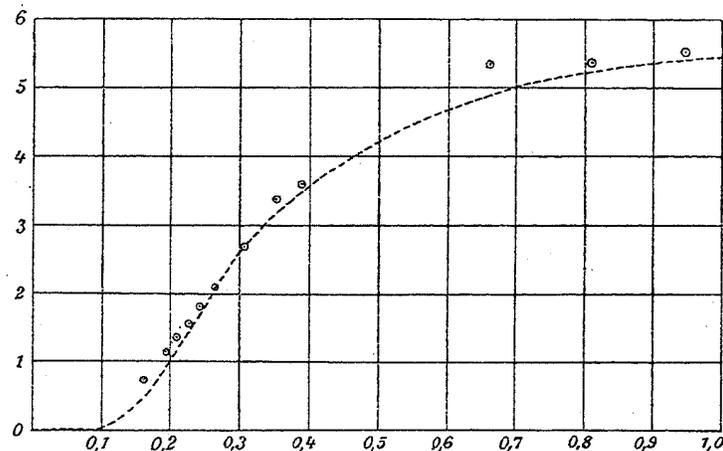


Figure 2.2: Plot of Molar Heat Capacity of Diamond from Einstein's Original 1907 paper. The fit is to the Einstein theory. The x -axis is $k_B T$ in units of $h\omega$ and the y axis is C in units of $\text{cal}/(\text{K}\cdot\text{mol})$. In these units, $3R \approx 5.96$.

of the heat capacity, but more importantly it told us something fundamental about quantum mechanics. Keep in mind that Einstein obtained this result 19 years before the Schrodinger equation was discovered!⁶

2.2 Debye's Calculation

Einstein's theory of specific heat was extremely successful, but still there were clear deviations from the predicted equation. Even in the plot in his first paper (Fig. 2.2 above) one can see that at low temperature the experimental data lies above the theoretical curve⁷. This result turns out to be rather important! In fact, it was known that at low temperatures most materials have a heat capacity that is proportional to T^3 (Metals also have a very small additional term proportional to T which we will discuss later in section 4.2. Magnetic materials may have other additional terms as well⁸. Nonmagnetic insulators have only the T^3 behavior). At any rate, Einstein's formula at low temperature is exponentially small in T , not agreeing at all with the actual experiments.

In 1912 Peter Debye⁹ discovered how to better treat the quantum mechanics of oscillations of atoms, and managed to explain the T^3 specific heat. Debye realized that oscillation of atoms is the same thing as sound, and sound is a wave, so it should be quantized the same way as Planck quantized light waves. Besides the fact that the speed of light is much faster than that of sound, there is only one minor difference between light and sound: for light, there are two polarizations for each \mathbf{k} whereas for sound, there are three modes for each \mathbf{k} (a longitudinal mode, where the atomic motion is in the same direction as \mathbf{k} and two transverse modes where the motion is perpendicular to \mathbf{k} . Light has only the transverse modes.). For simplicity of presentation here we will assume that the transverse and longitudinal modes have the same velocity, although in truth the longitudinal

⁶Einstein was a pretty smart guy.

⁷Although perhaps not obvious, this deviation turns out to be real, and not just experimental error.

⁸We will discuss magnetism in part VII.

⁹Peter Debye later won a Nobel prize in Chemistry for something completely different.

velocity is usually somewhat greater than the transverse velocity¹⁰.

We now repeat essentially what was Planck's calculation for light. This calculation should also look familiar from your statistical physics course. First, however, we need some preliminary information about waves:

2.2.1 About Periodic (Born-Von-Karman) Boundary Conditions

Many times in this course we will consider waves with periodic or "Born-Von-Karman" boundary conditions. It is easiest to describe this first in one dimension. Here, instead of having a one dimensional sample of length L with actual ends, we imagine that the two ends are connected together making the sample into a circle. The periodic boundary condition means that, any wave in this sample e^{ikr} is required to have the same value for a position r as it has for $r + L$ (we have gone all the way around the circle). This then restricts the possible values of k to be

$$k = \frac{2\pi n}{L}$$

for n an integer. If we are ever required to sum over all possible values of k , for large enough L we can replace the sum with an integral obtaining¹¹

$$\sum_k \rightarrow \frac{L}{2\pi} \int_{-\infty}^{\infty} dk$$

A way to understand this mapping is to note that the spacing between allowed points in k space is $2\pi/L$ so the integral $\int dk$ can be replaced by a sum over k points times the spacing between the points.

In three dimensions, the story is extremely similar. For a sample of size L^3 , we identify opposite ends of the sample (wrapping the sample up into a hypertorus!) so that if you go a distance L in any direction, you get back to where you started¹². As a result, our \mathbf{k} values can only take values

$$\mathbf{k} = \frac{2\pi}{L}(n_1, n_2, n_3)$$

for integer values of n_i , so here each \mathbf{k} point now occupies a volume of $(2\pi/L)^3$. Because of this discretization of values of \mathbf{k} , whenever we have a sum over all possible \mathbf{k} values we obtain

$$\sum_{\mathbf{k}} \rightarrow \frac{L^3}{(2\pi)^3} \int d\mathbf{k}$$

¹⁰We have also assumed the sound velocity to be the same in every direction, which need not be true in real materials. It is not too hard to include anisotropy into Debye's theory as well.

¹¹In your previous courses you may have used particle in a box boundary conditions where instead of plane waves $e^{i2\pi nr/L}$ you used particle in a box wavefunctions of the form $\sin(kn\pi r/L)$. This gives you instead

$$\sum_k \rightarrow \frac{L}{\pi} \int_0^{\infty} dk$$

which will inevitably result in the same physical answers as for the periodic boundary condition case. All calculations can be done either way, but periodic Born-Von-Karman boundary conditions are almost always simpler.

¹²Such boundary conditions are very popular in video games. It may also be possible that our universe has such boundary conditions — a notion known as the *doughnut universe*. Data collected by Cosmic Microwave Background Explorer (led by Nobel Laureates John Mather and George Smoot) and its successor the Wilkinson Microwave Anisotropy Probe appear consistent with this structure.

with the integral over all three dimensions of \mathbf{k} -space (this is what we mean by the bold \mathbf{dk}). One might think that wrapping the sample up into a hypertorus is very unnatural compared to considering a system with real boundary conditions. However, these boundary conditions tend to simplify calculations quite a bit and most physical quantities you might measure could be measured far from the boundaries of the sample anyway and would then be independent of what you do with the boundary conditions.

2.2.2 Debye's Calculation Following Planck

Debye decided that the oscillation modes were waves with frequencies $\omega(\mathbf{k}) = v|\mathbf{k}|$ with v the sound velocity — and for each \mathbf{k} there should be three possible oscillation modes, one for each direction of motion. Thus he wrote an expression entirely analogous to Einstein's expression (compare to Eq. 2.1)

$$\begin{aligned}\langle E \rangle &= 3 \sum_{\mathbf{k}} \hbar\omega(\mathbf{k}) \left(n_B(\beta\hbar\omega(\mathbf{k})) + \frac{1}{2} \right) \\ &= 3 \frac{L^3}{(2\pi)^3} \int \mathbf{dk} \hbar\omega(\mathbf{k}) \left(n_B(\beta\hbar\omega(\mathbf{k})) + \frac{1}{2} \right)\end{aligned}$$

Each excitation mode is a boson of frequency $\omega(\mathbf{k})$ and it is occupied on average $n_B(\beta\hbar\omega(\mathbf{k}))$ times.

By spherical symmetry, we may convert the three dimensional integral to a one dimensional integral

$$\int \mathbf{dk} \rightarrow 4\pi \int_0^\infty k^2 dk$$

(recall that $4\pi k^2$ is the area of the surface of a sphere¹³ of radius k) and we also use $k = \omega/v$ to obtain

$$\langle E \rangle = 3 \frac{4\pi L^3}{(2\pi)^3} \int_0^\infty \omega^2 d\omega (1/v^3) (\hbar\omega) \left(n_B(\beta\hbar\omega) + \frac{1}{2} \right)$$

It is convenient to replace $nL^3 = N$ where n is the density of atoms. We then obtain

$$\langle E \rangle = \int_0^\infty d\omega g(\omega) (\hbar\omega) \left(n_B(\beta\hbar\omega) + \frac{1}{2} \right) \quad (2.2)$$

where the *density of states* is given by

$$g(\omega) = N \left[\frac{12\pi\omega^2}{(2\pi)^3 n v^3} \right] = N \frac{9\omega^2}{\omega_d^3} \quad (2.3)$$

where

$$\omega_d^3 = 6\pi^2 n v^3 \quad (2.4)$$

This frequency will be known as the *Debye frequency*, and below we will see why we chose to define it this way with the factor of 9 removed.

The meaning of the density of states¹⁴ here is that the total number of oscillation modes with frequencies between ω and $\omega + d\omega$ is given by $g(\omega)d\omega$. Thus the interpretation of Eq. 2.2 is

¹³Or to be pedantic, $\int \mathbf{dk} \rightarrow \int_0^{2\pi} d\phi \int_0^\pi d\theta \sin\theta \int k^2 dk$ and performing the angular integrals gives 4π .

¹⁴We will encounter the concept of density of states many times, so it is a good idea to become comfortable with it!

simply that we should count how many modes there are per frequency (given by g) then multiply by the expected energy per mode (compare to Eq. 2.1) and finally we integrate over all frequencies. This result, Eq. 2.2, for the quantum energy of the sound waves is strikingly similar to Planck's result for the quantum energy of light waves, only we have replaced $2/c^3$ by $3/v^3$ (replacing the 2 light modes by 3 sound modes). The other change from Planck's classic result is the $+1/2$ that we obtain as the zero point energy of each oscillator¹⁵. At any rate, this zero point energy gives us a contribution which is temperature independent¹⁶. Since we are concerned with $C = \partial\langle E \rangle / \partial T$ this term will not contribute and we will separate it out. We thus obtain

$$\langle E \rangle = \frac{9N\hbar}{\omega_d^3} \int_0^\infty d\omega \frac{\omega^3}{e^{\beta\hbar\omega} - 1} + T \text{ independent constant}$$

by defining a variable $x = \beta\hbar\omega$ this becomes

$$\langle E \rangle = \frac{9N\hbar}{\omega_d^3(\beta\hbar)^4} \int_0^\infty dx \frac{x^3}{e^x - 1} + T \text{ independent constant}$$

The nasty integral just gives some number¹⁷ – in fact the number is $\pi^4/15$. Thus we obtain

$$\langle E \rangle = 9N \frac{(k_B T)^4 \pi^4}{(\hbar\omega_d)^3 15} + T \text{ independent constant}$$

Notice the similarity to Planck's derivation of the T^4 energy of photons. As a result, the heat capacity is

$$C = \frac{\partial\langle E \rangle}{\partial T} = Nk_B \frac{(k_B T)^3 12\pi^4}{(\hbar\omega_d)^3 5} \sim T^3$$

This correctly obtains the desired T^3 specific heat. Furthermore, the prefactor of T^3 can be calculated in terms of known quantities such as the sound velocity and the density of atoms. Note that the Debye frequency in this equation is sometimes replaced by a temperature

$$\hbar\omega_d = k_B T_{Debye}$$

known as the *Debye temperature*, so that this equation reads

$$C = \frac{\partial\langle E \rangle}{\partial T} = Nk_B \frac{(T)^3}{(T_{Debye})^3} \frac{12\pi^4}{5}$$

¹⁵Planck should have gotten this energy as well, but he didn't know about zero-point energy — in fact, since it was long before quantum mechanics was fully understood, Debye didn't actually have this term either.

¹⁶Temperature independent and also infinite. Handling infinities like this is something that gives mathematicians nightmares, but physicist do it happily when they know that the infinity is not really physical. We will see below in section 2.2.3 how this infinity gets properly cut off by the Debye Frequency.

¹⁷If you wanted to evaluate the nasty integral, the strategy is to reduce it to the famous Riemann zeta function. We start by writing

$$\int_0^\infty dx \frac{x^3}{e^x - 1} = \int_0^\infty dx \frac{x^3 e^{-x}}{1 - e^{-x}} = \int_0^\infty dx x^3 e^{-x} \sum_{n=0}^\infty e^{-nx} = \sum_{n=1}^\infty \int_0^\infty dx x^3 e^{-nx} = 3! \sum_{n=1}^\infty \frac{1}{n^4}$$

The resulting sum is a special case of the famous Riemann zeta function defined as $\zeta(p) = \sum_{n=1}^\infty n^{-p}$ where here we are concerned with the value of $\zeta(4)$. Since the zeta function is one of the most important functions in all of mathematics¹⁸, one can just look up its value on a table to find that $\zeta(4) = \pi^4/90$ thus giving us the above stated result that the nasty integral is $\pi^4/15$. However, in the unlikely event that you were stranded on a desert island and did not have access to a table, you could even evaluate this sum explicitly, which we do in the appendix to this chapter.

¹⁸One of the most important unproven conjectures in all of mathematics is known as the Riemann hypothesis and is concerned with determining for which values of p does $\zeta(p) = 0$. The hypothesis was written down in 1869 by Bernard Riemann (the same guy who invented Riemannian geometry, crucial to general relativity) and has defied proof ever since. The Clay Mathematics Institute has offered one million dollars for a successful proof.

2.2.3 Debye's "Interpolation"

Unfortunately, now Debye has a problem. In the expression derived above, the heat capacity is proportional to T^3 up to arbitrarily high temperature. We know however, that the heat capacity should level off to $3k_B N$ at high T . Debye understood that the problem with his approximation is that it allows an infinite number of sound wave modes — up to arbitrarily large \mathbf{k} . This would imply more sound wave modes than there are atoms in the entire system. Debye guessed (correctly) that really there should be only as many modes as there are degrees of freedom in the system. We will see in sections 8-12 below that this is an important general principle. To fix this problem, Debye decided to not consider sound waves above some maximum frequency ω_{cutoff} , with this frequency chosen such that there are exactly $3N$ sound wave modes in the system (3 dimensions of motion times N particles). We thus define ω_{cutoff} via

$$3N = \int_0^{\omega_{cutoff}} d\omega g(\omega) \quad (2.5)$$

We correspondingly rewrite Eq. 2.2 for the energy (dropping the zero point contribution) as

$$\langle E \rangle = \int_0^{\omega_{cutoff}} d\omega g(\omega) \hbar\omega n_B(\beta\hbar\omega) \quad (2.6)$$

Note that at very low temperature, this cutoff does not matter at all, since for large β the Bose factor n_B will very rapidly go to zero at frequencies well below the cutoff frequency anyway.

Let us now check that this cutoff gives us the correct high temperature limit. For high temperature

$$n_B(\beta\hbar\omega) = \frac{1}{e^{\beta\hbar\omega} - 1} \rightarrow \frac{k_B T}{\hbar\omega}$$

Thus in the high temperature limit, invoking Eqs. 2.5 and 2.6 we obtain

$$\langle E \rangle = k_B T \int_0^{\omega_{cutoff}} d\omega g(\omega) = 3k_B T N$$

yielding the Dulong-Petit high temperature heat capacity $C = \partial\langle E \rangle / \partial T = 3k_B N = 3k_B$ per atom. For completeness, let us now evaluate our cutoff frequency,

$$3N = \int_0^{\omega_{cutoff}} d\omega g(\omega) = 9N \int_0^{\omega_{cutoff}} d\omega \frac{\omega^2}{\omega_d^3} = 3N \frac{\omega_{cutoff}^3}{\omega_d^3}$$

we thus see that the correct cutoff frequency is exactly the Debye frequency ω_d . Note that $k = \omega_d/v = (6\pi^2 n)^{1/3}$ (from Eq. 2.4) is on the order of the inverse interatomic spacing of the solid.

More generally (in the neither high nor low temperature limit) one has to evaluate the integral 2.6, which cannot be done analytically. Nonetheless it can be done numerically and then can be compared to actual experimental data as shown in Fig. 2.3. It should be emphasized that the Debye theory makes predictions without any free parameters, as compared to the Einstein theory which had the unknown Einstein frequency ω as a free fitting parameter.

2.2.4 Some Shortcomings of the Debye Theory

While Debye's theory is remarkably successful, it does have a few shortcomings.

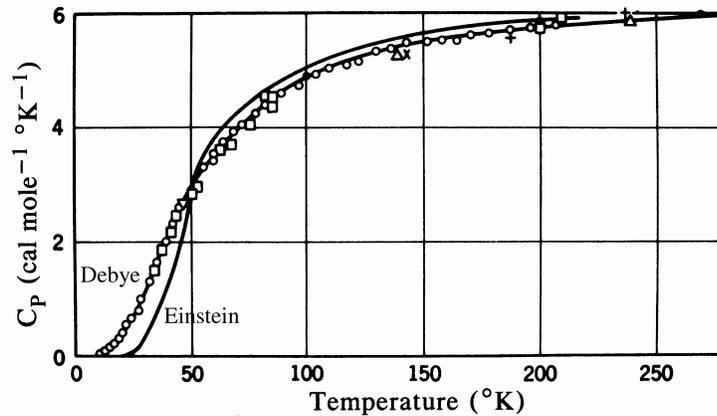


Figure 2.3: Plot of Heat Capacity of Silver. The y axis is C in units of $\text{cal}/(\text{K}\cdot\text{mol})$. In these units, $3R \approx 5.96$). Over the entire experimental range, the fit to the Debye theory is excellent. At low T it correctly recovers the T^3 dependence, and at high T it converges to the law of Dulong-Petit.

- The introduction of the cutoff seems very ad-hoc. This seems like a successful cheat rather than real physics
- We have assumed sound waves follow the law $\omega = vk$ even for very very large values of k (on the order of the inverse lattice spacing), whereas the entire idea of sound is a long wavelength idea, which doesn't seem to make sense for high enough frequency and short enough wavelength. At any rate, it is known that at high enough frequency the law $\omega = vk$ no longer holds.
- Experimentally, the Debye theory is very accurate, but it is not exact at intermediate temperatures.
- At very very low temperatures, metals have a term in the heat capacity that is proportional to T , so the overall heat capacity is $C = aT + bT^3$ and at low enough T the linear term will dominate¹⁹ You can't see this contribution on the plot Fig. 2.3 but at very low T it becomes evident.

Of these shortcomings, the first three can be handled more properly by treating the details of the crystal structure of materials accurately (which we will do much later in this course). The final issue requires us to carefully study the behavior of electrons in metals to discover the origin of this linear T term (see section 4.2 below).

Nonetheless, despite these problems, Debye's theory was a substantial improvement over Einstein's²⁰,

¹⁹In magnetic materials there may be still other contributions to the heat capacity reflecting the energy stored in magnetic degrees of freedom. See part VII below.

²⁰Debye was pretty smart too... even though he was a chemist.

2.3 Summary of Specific Heat of Solids

- (Much of the) Heat capacity (specific heat) of materials is due to atomic vibrations.
- Boltzmann and Einstein models consider these vibrations as N simple harmonic oscillators.
- Boltzmann classical analysis obtains law of Dulong-Petit $C = 3Nk_B = 3R$.
- Einstein quantum analysis shows that at temperatures below the oscillator frequency, degrees of freedom freeze out, and heat capacity drops exponentially. Einstein frequency is a fitting parameter.
- Debye Model treats oscillations as sound waves. No fitting parameters.
 - $\omega = v|k|$, similar to light (but three polarizations not two)
 - quantization similar to Planck quantization of light
 - Maximum frequency cutoff ($\hbar\omega_{Debye} = k_B T_{Debye}$) necessary to obtain a total of only $3N$ degrees of freedom
 - obtains Dulong-Petit at high T and $C \sim T^3$ at low T .
- Metals have an additional (albeit small) linear T term in the heat capacity which we will discuss later.

References

Almost every book covers the material introduced in this chapter, but frequently it is done late in the book only after the idea of phonons is introduced. We will get to phonons in chapter 8. Before we get there the following references cover this material without discussion of phonons:

- Goodstein sections 3.1 and 3.2
- Rosenberg sections 5.1 through 5.13 (good problems included)
- Burns sections 11.3 through 11.5 (good problems included)

Once we get to phonons, we can look back at this material again. Discussions are then given also by

- Dove section 9.1 and 9.2
- Ashcroft and Mermin chapter 23
- Hook and Hall section 2.6
- Kittel beginning of chapter 5

2.4 Appendix to this Chapter: $\zeta(4)$

The Riemann zeta function as mentioned above is defined as

$$\zeta(p) = \sum_{n=1}^{\infty} n^{-p}.$$

This function occurs frequently in physics, not only in the Debye theory of solids, but also in the Sommerfeld theory of electrons in metals (see chapter 4 below), as well as in the study of Bose condensation. As mentioned above in footnote 18 of this chapter, it is also an extremely important quantity to mathematicians.

In this appendix we are concerned with the value of $\zeta(4)$. To evaluate this we write a Fourier series for the function x^2 on the interval $[-\pi, \pi]$. The series is given by

$$x^2 = \frac{a_0}{2} + \sum_{n>0} a_n \cos(nx)$$

with coefficients given by

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} dx x^2 \cos(nx)$$

These can be calculated straightforwardly to give

$$a_n = \begin{cases} 2\pi^2/3 & n = 0 \\ 4(-1)^n/n^2 & n > 0 \end{cases}$$

We now calculate an integral in two different ways. First we can directly evaluate

$$\int_{-\pi}^{\pi} dx (x^2)^2 = \frac{2\pi^5}{5}$$

On the other hand using the above Fourier decomposition of x^2 we can write the same integral as

$$\begin{aligned} \int_{-\pi}^{\pi} dx (x^2)^2 &= \int_{-\pi}^{\pi} dx \left(\frac{a_0}{2} + \sum_{n>0} a_n \cos(nx) \right) \left(\frac{a_0}{2} + \sum_{m>0} a_m \cos(mx) \right) \\ &= \int_{-\pi}^{\pi} dx \left(\frac{a_0}{2} \right)^2 + \int_{-\pi}^{\pi} dx \sum_{n>0} (a_n \cos(nx)) \end{aligned}$$

where we have used the orthogonality of Fourier modes to eliminate cross terms in the product. We can do these integrals to obtain

$$\int_{-\pi}^{\pi} dx (x^2)^2 = \pi \left(\frac{a_0^2}{2} + \sum_{n>0} a_n^2 \right) = \frac{2\pi^5}{9} + 16\pi\zeta(4)$$

Setting this expression to $2\pi^5/5$ gives us the result $\zeta(4) = \pi^4/90$.

Chapter 3

Electrons in Metals: Drude Theory

The fundamental characteristic of a metal is that it conducts electricity. At some level the reason for this conduction boils down to the fact that electrons are mobile in these materials. In later chapters we will be concerned with the question of why electrons are mobile in some materials but not in others, being that all materials have electrons in them! For now, we take as given that there are mobile electrons and we would like to understand their properties.

J.J. Thomson's 1896 discovery of the electron ("corpuscles of charge" that could be pulled out of metal) raised the question of how these charge carriers might move within the metal. In 1900 Paul Drude¹ realized that he could apply Boltzmann's kinetic theory of gases to understanding electron motion within metals. This theory was remarkably successful, providing a first understanding of metallic conduction.²

Having studied the kinetic theory of gases, Drude theory should be very easy to understand. We will make three assumptions about the motion of electrons

1. Electrons have a scattering time τ . The probability of scattering within a time interval dt is dt/τ .
2. Once a scattering event occurs, we assume the electron returns to momentum $\mathbf{p} = 0$.
3. In between scattering events, the electrons, which are charge $-e$ particles, respond to the externally applied electric field \mathbf{E} and magnetic field \mathbf{B} .

The first two of these assumptions are exactly those made in the kinetic theory of gases³. The third assumption is just a logical generalization to account for the fact that, unlike gases molecules,

¹pronounced roughly "Drood-a"

²Sadly, neither Boltzmann nor Drude lived to see how much influence this theory really had — in unrelated tragic events, both of them committed suicide in 1906. Boltzmann's famous student, Ehrenfest, also committed suicide some years later. Why so many highly successful statistical physicists took their own lives is a bit of a mystery.

³Ideally we would do a better job with our representation of the scattering of particles. Every collision should consider two particles having initial momenta $\mathbf{p}_1^{initial}$ and $\mathbf{p}_2^{initial}$ and then scattering to final momenta \mathbf{p}_1^{final} and \mathbf{p}_2^{final} so as to conserve both energy and momentum. Unfortunately, keeping track of things so carefully makes the problem extremely difficult to solve. Assumption 1 is not so crazy as an approximation being that there really is a typical time between scattering events in a gas. Assumption 2 is a bit more questionable, but *on average* the final

electrons are charged and must therefore respond to electromagnetic fields.

We consider an electron with momentum \mathbf{p} at time t and we ask what momentum it will have at time $t + dt$. There are two terms in the answer, there is a probability dt/τ that it will scatter to momentum zero. If it does not scatter to momentum zero (with probability $1 - dt/\tau$) it simply accelerates as dictated by its usual equations of motion $d\mathbf{p}/dt = \mathbf{F}$. Putting the two terms together we have

$$\langle \mathbf{p}(t + dt) \rangle = \left(1 - \frac{dt}{\tau}\right) (\mathbf{p}(t) + \mathbf{F}dt) + \mathbf{0} dt/\tau$$

or⁴

$$\frac{d\mathbf{p}}{dt} = \mathbf{F} - \frac{\mathbf{p}}{\tau} \quad (3.1)$$

where here the force \mathbf{F} on the electron is just the Lorentz force

$$\mathbf{F} = -e(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

One can think of the scattering term $-\mathbf{p}/\tau$ as just a drag force on the electron. Note that in the absence of any externally applied field the solution to this differential equation is just an exponentially decaying momentum

$$\mathbf{p}(t) = \mathbf{p}_{initial} e^{-t/\tau}$$

which is what we should expect for particles that lose momentum by scattering.

3.1 Electrons in Fields

3.1.1 Electrons in an Electric Field

Let us start by considering the case where the electric field is nonzero but the magnetic field is zero. Our equation of motion is then

$$\frac{d\mathbf{p}}{dt} = -e\mathbf{E} - \frac{\mathbf{p}}{\tau}$$

In steady state, $d\mathbf{p}/dt = 0$ so we have

$$m\mathbf{v} = \mathbf{p} = -e\tau\mathbf{E}$$

with m the mass of the electron and \mathbf{v} its velocity.

Now, if there is a density n of electrons in the metal each with charge $-e$, and they are all moving at velocity \mathbf{v} , then the electrical current is given by

$$\mathbf{j} = -en\mathbf{v} = \frac{e^2\tau n}{m}\mathbf{E}$$

momentum after a scattering event is indeed zero (if you average momentum as a vector). However, obviously it is not correct that every particle has zero kinetic energy after a scattering event. This is a defect of the approach.

⁴Here we really mean $\langle \mathbf{p} \rangle$ when we write \mathbf{p} . Since our scattering is probabilistic, we should view all quantities (such as the momentum) as being an expectation over these random events. A more detailed theory would keep track of the entire distribution of momenta rather than just the average momentum. Keeping track of distributions in this way leads one to the Boltzmann Transport Equation, which we will not discuss.

or in other words, the conductivity of the metal, defined via $\mathbf{j} = \sigma \mathbf{E}$ is given by⁵

$$\sigma = \frac{e^2 \tau n}{m} \quad (3.2)$$

By measuring the conductivity of the metal (assuming we know both the charge and mass of the electron) we can determine the product of the density and scattering time of the electron.

3.1.2 Electrons in Electric and Magnetic Fields

Let us continue on to see what other predictions come from Drude theory. Consider the transport equation 3.1 for a system in both an electric and a magnetic field. We now have

$$\frac{d\mathbf{p}}{dt} = -e(\mathbf{E} + \mathbf{v} \times \mathbf{B}) - \mathbf{p}/\tau$$

Again setting this to zero in steady state, and using $\mathbf{p} = m\mathbf{v}$ and $\mathbf{j} = -nev$, we obtain an equation for the steady state current

$$0 = -e\mathbf{E} + \frac{\mathbf{j} \times \mathbf{B}}{n} + \frac{m}{ne\tau}\mathbf{j}$$

or

$$\mathbf{E} = \left(\frac{1}{ne}\mathbf{j} \times \mathbf{B} + \frac{m}{ne^2\tau}\mathbf{j} \right)$$

We now define the 3 by 3 resistivity matrix $\underline{\rho}$ which relates the current vector to the electric field vector

$$\mathbf{E} = \underline{\rho}\mathbf{j}$$

such that the components of this matrix are given by

$$\rho_{xx} = \rho_{yy} = \rho_{zz} = \frac{m}{ne^2\tau}$$

and if we imagine \mathbf{B} oriented in the \hat{z} direction, then

$$\rho_{xy} = -\rho_{yx} = \frac{B}{ne}$$

and all other components of $\underline{\rho}$ are zero. This off-diagonal term in the resistivity is known as the *Hall resistivity*, named after Edwin Hall who discovered in 1879 that when a magnetic field is applied perpendicular to a current flow, a voltage can be measured perpendicular to both current and magnetic field (See Fig. 3.1.2). As a homework problem you might consider a further generalization of Drude theory to finite frequency conductivity, where it gives some interesting (and frequently accurate) predictions.

The Hall coefficient R_H is defined as

$$R_H = \frac{\rho_{yx}}{|B|}$$

which in the Drude theory is given by

$$R_H = \frac{-1}{ne}$$

⁵A related quantity is the *mobility*, defined by $\mathbf{v} = \mu\mathbf{F}$, which is given in Drude theory by $\mu = e\tau/m$. We will discuss mobility further in section 16.1.1 below.

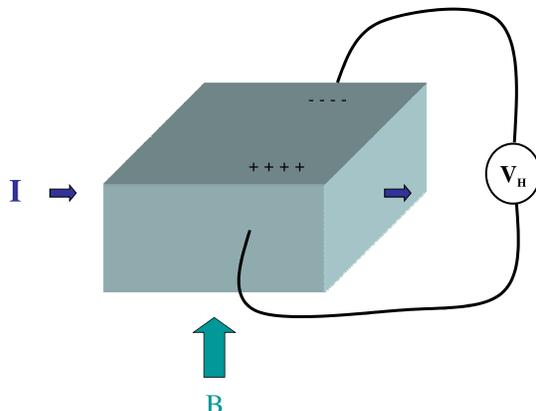


Figure 3.1: Edwin Hall's 1879 experiment. The voltage measured perpendicular to both the magnetic field and the current is known as the Hall voltage which is proportional to B and inversely proportional to the electron density (at least in Drude theory).

This then allows us to measure the density of electrons in a metal.

Aside: One can also consider turning this experiment on its head. If you know the density of electrons in your sample you can use a Hall measurement to determine the magnetic field. This is known as a Hall sensor. Since it is hard to measure small voltages, Hall sensors typically use materials, such as semiconductors, where the density of electrons is low so R_H and hence the resulting voltage is large.

Let us then calculate $n = -1/(eR_H)$ for various metals and divide it by the density of atoms. This should give us the number of free electrons per atom. Later on we will see that it is frequently not so hard to estimate the number of electrons in a system. A short description is that electrons bound in the core shells of the atoms are never free to travel throughout the crystal, whereas the electrons in the outer shell may be free (we will discuss later when these electrons are free and when they are not). The number of electrons in the outermost shell is known as the *valence* of the atom.

Material	$(-1/[eR_H])/ [\text{density of atoms}]$ In Drude theory this should give the number of free electrons per atom which is the valence	Valence
Li	.8	1
Na	1.2	1
K	1.1	1
Cu	1.5	1 (usually)
Be	-0.2 (but anisotropic)	2
Mg	-0.4	2

Table 3.1: Comparison of the valence of various atoms to the measured number of free electrons per atom (measured via the Hall resistivity and the atomic density).

We see from table 3.1 that for many metals this Drude theory analysis seems to make sense — the “valence” of lithium, sodium, and potassium (Li, Na, and K) are all one which agrees roughly with the measured number of electrons per atom. The effective valence of copper (Cu) is also one, so it is not surprising either. However, something has clearly gone seriously wrong for Be and Mg. In this case, the sign of the Hall coefficient has come out incorrect. From this result, one might conclude that the charge carrier for beryllium and magnesium (Be and Mg) have the opposite charge from that of the electron! We will see below in section 16.1.1 that this is indeed true and is a result of the so-called band structure of these materials. However, for many metals, simple Drude theory gives quite reasonable results. We will see in chapter 16 below that Drude theory is particularly good for describing semiconductors.

If we believe the Hall effect measurement of the density of electrons in metals, using Eq. 3.2 we can then extract a scattering time from the expression for the conductivity. The Drude scattering time comes out to be in the range of $\tau \approx 10^{-14}$ seconds for most metals near room temperature.

3.2 Thermal Transport

Drude was brave enough to attempt to further calculate the thermal conductivity κ due to mobile electrons⁶ using Boltzmann’s kinetic theory. Without rehashing the derivation, this result should look familiar to you from your previous encounters with the kinetic theory of gas

$$\kappa = \frac{1}{3} n c_v \langle v \rangle \lambda$$

where c_v is the heat capacity per particle, $\langle v \rangle$ is the average thermal velocity and $\lambda = \langle v \rangle \tau$ is the scattering length. For a conventional gas the heat capacity per particle is

$$c_v = \frac{3}{2} k_B$$

and

$$\langle v \rangle = \sqrt{\frac{8k_B T}{\pi m}}$$

Assuming this all holds true for electrons, we obtain

$$\kappa = \frac{4}{\pi} \frac{n \tau k_B^2 T}{m}$$

While this quantity still has the unknown parameter τ in it, it is the same quantity that occurs in the electrical conductivity (Eq. 3.2). Thus we may look at the ratio of thermal conductivity to

⁶In any experiment there will also be some amount of thermal conductivity from structural vibrations of the material as well — so called phonon thermal conductivity. (We will meet phonons in chapter 8 below). However, for most metals, the thermal conductivity is mainly due to electron motion and not from vibrations.

electrical conductivity, known as the Lorenz number^{7,8}

$$L = \frac{\kappa}{T\sigma} = \frac{4}{\pi} \left(\frac{k_B}{e} \right)^2 \approx 0.94 \times 10^{-8} \text{ WattOhm/K}^2$$

A slightly different prediction is obtained by realizing that we have used $\langle v \rangle^2$ in our calculation, whereas perhaps we might have instead used $\langle v^2 \rangle$ which would have then given us instead

$$L = \frac{\kappa}{T\sigma} = \frac{3}{2} \left(\frac{k_B}{e} \right)^2 \approx 1.11 \times 10^{-8} \text{ WattOhm/K}^2$$

This result was viewed as a huge success, being that it was known for almost half a century that almost all metals have roughly the same value of this ratio, a fact known as the Wiedemann-Franz law. In fact the value predicted for this ratio is only a bit lower than that measured experimentally (See table 3.2).

Material	$L \times 10^8$ (WattOhm/K ²)
Li	2.22
Na	2.12
Cu	2.20
Fe	2.61
Bi	3.53
Mg	2.14
Drude Prediction	0.98-1.11

Table 3.2: Lorenz Numbers $\kappa/(T\sigma)$ for Various Metals

So the result appears to be off by about a factor of 2, but still that is very good, considering that before Drude no one had any idea why this ratio should be a constant at all!

In retrospect we now realize that this calculation is completely incorrect (despite its successful result). The reason we know there is a problem is because we do not actually measure a specific heat of $C_v = \frac{3}{2}k_B$ per electron in metals (for certain systems where the density of electrons is very low, we do in fact measure this much specific heat, but not in metals). In fact, in most metals we measure only a vibrational (Debye) specific heat, plus a very small term linear in T at low temperatures. So why does this calculation give such a good result? It turns out (and we will see later below) that we have made two mistakes that roughly cancel each other. We have used a specific heat that is way too large, but we have also used a velocity that is way too small. We will see later that both of these mistakes are due to Fermi statistics of the electron (which we have so far ignored) and the Pauli exclusion principle.

We can see the problem much more clearly in some other quantities. The so-called *Peltier effect* is the fact that running electrical current through a material also transports heat. The

⁷This is named after Ludvig Lorenz, not Hendrik Lorentz who is famous for the Lorentz force and Lorentz contraction. However, just to confuse matters, the two of them worked on similar topics and there is even a Lorentz-Lorenz equation

⁸The dimensions here might look a bit funny, but κ , the thermal conductivity is measured in Watt/K and σ is measured in 1/Ohm. To see that WattOhm/K² is the same as $(k_B/e)^2$ note that k_B is J/K and e is Coulomb (C). So we need to show that (J/C)² is WattOhm

$$(J/C)^2 = (J/\text{sec})(J/C)(1/(C/\text{sec})) = \text{Watt Volt/Amp} = \text{WattOhm}$$

so-called Peltier coefficient Π is defined by

$$\mathbf{j}^q = \Pi \mathbf{j}$$

where \mathbf{j}^q is the heat current, and \mathbf{j} is the electrical current.

Aside: The Peltier effect is used for thermoelectric refrigeration devices. By running electricity through a thermoelectric material, you can force heat to be transported through that material. You can thus transport heat away from one object and towards another. A good thermoelectric device has a high Peltier coefficient, but must also have a low resistivity, because running a current through an material with resistivity R will result in power I^2R being dissipated thus heating it up.

In kinetic theory the thermal current is

$$\mathbf{j}^q = \frac{1}{3}(c_v T)n\mathbf{v} \quad (3.3)$$

here $c_v T$ is the heat carried by one particle (with $c_v = 3k_B/2$ the heat capacity per particle) and n is the density of particles (and $1/3$ is the geometric factor that is probably approximate anyway). Similarly the electrical current is

$$\mathbf{j} = -en\mathbf{v}$$

Thus the Peltier coefficient is

$$\Pi = \frac{-c_v T}{3e} = \frac{-k_B T}{2e} \quad (3.4)$$

so the ratio (known as thermopower, or Seebeck coefficient) $S = \Pi/T$ is given by

$$S = \frac{\Pi}{T} = \frac{-k_B}{2e} = -4.3 \times 10^{-4} \text{V/K} \quad (3.5)$$

in Drude theory. For most metals the actual value of this ratio is roughly 100 times smaller! This is a reflection of the fact that we have used $c_v = 3k_B/2$ whereas the actual specific heat per particle is much much lower (which we will understand in the next section when we consider Fermi statistics more carefully).

3.3 Summary of Drude Theory

- Based on kinetic theory of gases.
- Assumes some scattering time τ , resulting in a conductivity $\sigma = ne^2\tau/m$.
- Hall coefficient measures density of electrons.
- Successes
 - Wiedemann-Franz ratio $\kappa/(\sigma T)$ comes out close to right for most materials
 - Many other transport properties predicted correctly (ex, conductivity at finite frequency)
 - Hall coefficient measurement of the density seems reasonable for many metals.
- Failures
 - Hall coefficient frequently is measured to have the wrong sign, indicating a charge carrier with charge opposite to that of the electron

- There is no $3k_B/2$ heat capacity per particle measured for electrons in metals. This then makes the Peltier coefficient come out wrong by a factor of 100.

The latter of the two shortcomings will be addressed in the next section, whereas the former of the two will be addressed in chapter 16 below where we discuss band theory.

Despite the shortcomings of Drude theory, it nonetheless was the only theory of metallic conductivity for a quarter of a century (until the Sommerfeld theory improved it), and it remains quite useful today (Particularly for semiconductors and other systems with low densities of electrons. See chapter 16).

References

- Ashcroft and Mermin, chapter 1
- Burns, chapter 9 part A
- Singleton, section 1.1–1.4
- Hook and Hall section 3.3 sort-of

Actually, Hook and Hall are aimed mainly at Free electron (Sommerfeld) theory (our next chapter), but they end up doing Drude theory anyway (they don't use the word "Drude").

Chapter 4

More Electrons in Metals: Sommerfeld (Free Electron) Theory

In 1925 Pauli discovered the exclusion principle, that no two electrons may be in the exact same state. In 1926, Fermi and Dirac separately derived what we now call Fermi-Dirac statistics¹ Upon learning about Fermi Statistics, Sommerfeld² realized that Drude's theory of metals could easily be generalized to incorporate Fermi statistics, which is what we shall presently do.

¹All three, Pauli, Fermi, and Dirac, won Nobel prizes in the next few years — but you probably knew that already.

²Sommerfeld never won a Nobel prize, although he was nominated for it 81 times — more than any other physicist. He also was a research advisor for more Nobel Laureates than anyone else in history (6: Heisenberg, Pauli, Debye, Bethe, who were his PhD students and Pauling, Rabi who were postdoctoral researchers with him. He also was the first research advisor for Rudolf Peierls for whom the theory building at Oxford is named, although Peierls eventually finished his PhD as a student of Pauli.)

4.1 Basic Fermi-Dirac Statistics

Given a system of free³ electrons with chemical potential⁴ μ the probability of an eigenstate of energy E being occupied is given by the Fermi factor⁵ (See Fig. 4.1)

$$n_F(\beta(E - \mu)) = \frac{1}{e^{\beta(E - \mu)} + 1} \quad (4.1)$$

At low temperature the Fermi function becomes a step function (states below the chemical potential are filled, those above the chemical potential are empty), whereas at higher temperatures the step function becomes more smeared out.

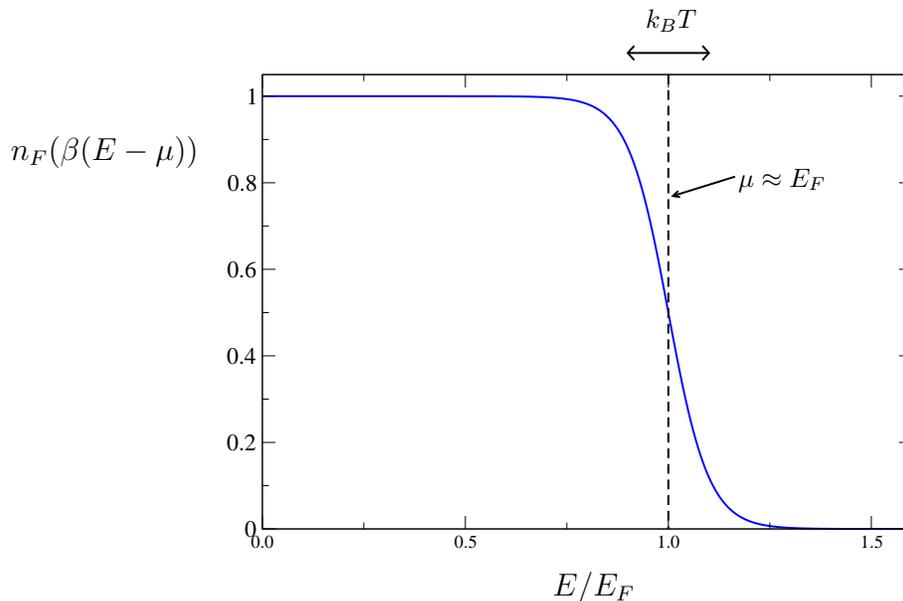


Figure 4.1: The Fermi Distribution for $k_B T \ll E_F$.

We will consider the electrons to be in a box of size $V = L^3$ and, as with our above discussion of sound waves, it is easiest to imagine that the box has periodic boundary conditions (See section 2.2.1). The plane wavefunctions are of the form $e^{i\mathbf{k}\cdot\mathbf{r}}$ where \mathbf{k} must take value $(2\pi/L)(n_1, n_2, n_3)$ with n_i integers due to the boundary conditions. These plane waves have corresponding energies

³Here “free” means that they do not interact with each other, with the background crystal lattice, with impurities, or with anything else for that matter.

⁴In case you did not properly learn about chemical potential in your statistical physics course, it can be *defined* via Eq. 4.1, by saying that μ is whatever constant needs to be inserted into this equation to make it true. It can also be defined as an appropriate thermodynamical derivative such as $\mu = \partial U / \partial N|_{V,S}$ with U the total energy and N the number of particles or $\mu = \partial G / \partial N|_{T,P}$ with G the Gibbs potential. However, such a definition can be tricky if one worries about the discreteness of the particle number — since N must be an integer, the derivative may not be well defined. As a result the definition in terms of Eq. 4.1 is frequently best (i.e. we are treating μ as a Lagrange multiplier).

⁵When we say that there are a particular set of N orbitals occupied by electrons, we really mean that the overall wavefunction of the system is an antisymmetric function $\Psi(1, \dots, N)$ which can be expressed as a Slater determinant of N particle coordinates occupying the N orbitals. We will never need to actually write out such Slater determinant wavefunctions except in Appendix 22.4 which is too advanced for any reasonable exam.

$$\epsilon(\mathbf{k}) = \frac{\hbar^2 |\mathbf{k}|^2}{2m} \quad (4.2)$$

with m the electron mass. Thus the total number of electrons in the system is given by

$$N = 2 \sum_{\mathbf{k}} n_F(\beta(\epsilon(\mathbf{k}) - \mu)) = 2 \frac{V}{(2\pi)^3} \int d\mathbf{k} n_F(\beta(\epsilon(\mathbf{k}) - \mu)) \quad (4.3)$$

where the prefactor of 2 accounts for the two possible spin states for each possible wavevector \mathbf{k} . In fact, in a metal, N will usually be given to us, and this equation will define the chemical potential as a function of temperature.

We now define a useful concept:

Definition 4.1.1. The *Fermi Energy*, E_F is the chemical potential at temperature $T = 0$.

This is also sometimes called the *Fermi Level*. The states that are filled at $T = 0$ are sometimes called the *Fermi Sea*. Frequently one also defines a *Fermi Temperature* $T_F = E_F/k_B$, and also the *Fermi Wavevector* k_F defined via

$$E_F = \frac{\hbar^2 k_F^2}{2m} \quad (4.4)$$

and correspondingly a *Fermi momentum* $p_F = \hbar k_F$ and a *Fermi velocity*⁶

$$v_F = \hbar k_F / m \quad (4.5)$$

Aside: Frequently people think of the Fermi Energy as the energy of the most energetic occupied electron state in system. While this is correct in the case where you are filling a continuum of states, it can also lead you to errors in cases where the energy eigenstates are discrete (see the related footnote 4 of this chapter), or more specifically when there is a gap between the most energetic occupied electron state in the system, and the least energetic unoccupied electron state. More correctly the Fermi energy, i.e., the chemical potential at $T = 0$, will be half-way between the most energetic occupied electron state, and the least energetic unoccupied electron state.

Let us now calculate the Fermi energy in a (three dimensional) metal with N electrons in it. At $T = 0$ the Fermi function (Eq. 4.1) becomes a step function (which we write as Θ . I.e., $\Theta(x) = 1$ for $x > 0$ and $= 0$ for $x < 0$), so that Eq. 4.3 becomes

$$N = 2 \frac{V}{(2\pi)^3} \int d\mathbf{k} \Theta(E_F - \epsilon(\mathbf{k})) = 2 \frac{V}{(2\pi)^3} \int^{|\mathbf{k}| < k_F} d\mathbf{k} = 2 \frac{V}{(2\pi)^3} \left(\frac{4}{3} \pi k_F^3 \right) \quad (4.6)$$

where in the last step we have use the fact that the volume of a ball is $4\pi/3$ times the cube of the radius, and at $T = 0$ the electrons fill a ball up to radius k_F . The surface of this ball, a sphere (the “Fermi sphere”) of radius k_F is known as the *Fermi Surface* — a term more generally defined as the surface dividing filled from unfilled states at zero temperature.

Using the fact that the density $n = N/V$ we can rearrange Eq. 4.6 to give

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

⁶Yes, Fermi got his name attached to many things. To help spread the credit around I’ve called this section “Basic Fermi-Dirac Statistics” instead of just “Basic Fermi Statistics”.

and correspondingly

$$E_F = \frac{\hbar^2(3\pi^2n)^{2/3}}{2m} \quad (4.7)$$

Since we know roughly how many free electrons there are in a metal (say, one per atom for monovalent metals such as sodium or copper), we can estimate the Fermi energy, which, say for copper, turns out to be on the order of 7 eV, corresponding to a Fermi temperature of about 80,000 K. (!). This amazingly high energy scale is a result of Fermi statistics and the very high density of electrons in metals. It is crucial to remember that for all metals, $T_F \gg T$ for any temperature anywhere near room temperature. In fact metals melt (and even vaporize!) at temperatures far far below their Fermi temperatures.

Similarly, one can calculate the Fermi velocity, which, for a typical metal such as copper, may be as large as 1% the speed of light! Again, this enormous velocity stems from the Pauli exclusion principle — all the lower momentum states are simply filled, so if the density of electrons is very high, the velocities will be very high as well.

With a Fermi energy that is so large, and therefore a Fermi sea that is very deep, any (not insanely large) temperature can only make excitations of electrons that are already very close to the Fermi surface (i.e., they can jump from just below the Fermi surface to just above with only a small energy increase). The electrons deep within the Fermi sea, near $\mathbf{k} = \mathbf{0}$, cannot be moved by any reasonably low energy perturbation simply because there are no available unfilled states for them to move to unless they absorb a very large amount of energy.

4.2 Electronic Heat Capacity

We now turn to examine the heat capacity of electrons in a metal. Analogous to Eq. 4.3, the total energy of our system of electrons is given now by

$$E_{total} = \frac{2V}{(2\pi)^3} \int \mathbf{dk} \epsilon(\mathbf{k}) n_F(\beta(\epsilon(\mathbf{k}) - \mu)) = \frac{2V}{(2\pi)^3} \int_0^\infty 4\pi k^2 dk \epsilon(\mathbf{k}) n_F(\beta(\epsilon(\mathbf{k}) - \mu))$$

where the chemical potential is defined as above by

$$N = \frac{2V}{(2\pi)^3} \int \mathbf{dk} n_F(\beta(\epsilon(\mathbf{k}) - \mu)) = \frac{2V}{(2\pi)^3} \int_0^\infty 4\pi k^2 dk n_F(\beta(\epsilon(\mathbf{k}) - \mu))$$

(Here we have changed to spherical coordinates to obtain a one dimensional integral and a factor of $4\pi k^2$ out front).

It is convenient to replace k in this equation by the energy E by using Eq. 4.2 or equivalently

$$k = \sqrt{\frac{2\epsilon m}{\hbar^2}}$$

we then have

$$dk = \sqrt{\frac{m}{2\epsilon\hbar^2}} d\epsilon$$

We can then rewrite these expressions as

$$E_{total} = V \int_0^\infty d\epsilon \epsilon g(\epsilon) n_F(\beta(\epsilon - \mu)) \quad (4.8)$$

$$N = V \int_0^\infty d\epsilon g(\epsilon) n_F(\beta(\epsilon - \mu)) \quad (4.9)$$

where

$$g(\epsilon)d\epsilon = \frac{2}{(2\pi)^3}4\pi k^2 dk = \frac{2}{(2\pi)^3}4\pi \left(\frac{2\epsilon m}{\hbar^2}\right) \sqrt{\frac{m}{2\epsilon\hbar^2}} d\epsilon = \frac{(2m)^{3/2}}{2\pi^2\hbar^3}\epsilon^{1/2}d\epsilon \quad (4.10)$$

is the *density of states per unit volume*. The definition⁷ of this quantity is such that $g(\epsilon)d\epsilon$ is the total number of eigenstates (including both spin states) with energies between ϵ and $\epsilon + d\epsilon$.

From Eq. 4.7 we can simply derive $(2m)^{3/2}/\hbar^3 = 3\pi^2 n/E_F^{3/2}$, thus we can simplify the density of states expression to

$$g(\epsilon) = \frac{3n}{2E_F} \left(\frac{\epsilon}{E_F}\right)^{1/2} \quad (4.11)$$

which is a fair bit simpler. Note that the density of states has dimensions of a density (an inverse volume) divided by an energy. It is clear that this is the dimensions it must have given Eq. 4.9 for example.

Note that the expression Eq. 4.9 should be thought of as defining the chemical potential given the number of electrons in the system and the temperature. Once the chemical potential is fixed, then Eq. 4.8 gives us the total kinetic energy of the system. Differentiating that quantity would give us the heat capacity. Unfortunately there is no way to do this analytically in all generality. However, we can use to our advantage that $T \ll T_F$ for any reasonable temperature, so that the Fermi factors n_F are close to a step function. Such an expansion was first used by Sommerfeld, but it is algebraically rather complicated⁸ (See Ashcroft and Mermin chapter 2 to see how it is done in detail). However, it is not hard to make an estimate of what such a calculation must give — which we shall now do.

When $T = 0$ the Fermi function is a step function and the chemical potential is (by definition) the Fermi energy. For small T , the step function is smeared out as we see in Fig. 4.1. Note, however, that in this smearing the number of states that are removed from below the chemical potential is almost exactly the same as the number of states that are added above the chemical potential⁹. Thus, for small T , one does not have to move the chemical potential much from the Fermi energy in order to keep the number of particles fixed in Eq. 4.9. We conclude that $\mu \approx E_F$ for any low temperature. (In fact, in more detail we find that $\mu(T) = E_F + \mathcal{O}(T/T_F)^2$, see Ashcroft and Mermin chapter 2).

Thus we can focus on Eq. 4.8 with the assumption that $\mu = E_F$. At $T = 0$ let us call the kinetic energy¹⁰ of the system $E(T = 0)$. At finite temperature, instead of a step function in Eq. 4.8 the step is smeared out as in Fig. 4.1. We see in the figure that only electrons within an energy range of roughly $k_B T$ of the Fermi surface can be excited — in general they are excited above the Fermi surface by an energy of about $k_B T$. Thus we can approximately write

$$E(T) = E(T = 0) + (\gamma/2)[Vg(E_F)(k_B T)](k_B T) + \dots$$

Here $Vg(E_F)$ is the density of states near the Fermi surface (Recall g is the density of states per unit volume), so the number of particles close enough to the Fermi surface to be excited is $Vg(E_F)(k_B T)$, and the final factor of $(k_B T)$ is roughly the amount of energy that each one gets

⁷Compare the physical meaning of this definition to that of the density of states for sound waves given in Eq. 2.3 above.

⁸Such a calculation requires, among other things, the evaluation of some very nasty integrals which turn out to be related to the Riemann Zeta function (see section 2.4 above).

⁹Since the Fermi function has a precise symmetry around μ given by $n_F(\beta(E - \mu)) = 1 - n_F(\beta(\mu - E))$, this equivalence of states removed from below the chemical potential and states inserted above would be an exact statement if the density of states in Eq. 4.9 were independent of energy.

¹⁰In fact $E(T = 0) = (3/5)NE_F$, which is not too hard to show. Try showing it!

excited by. Here γ is some constant which we cannot get right by such an approximate argument (but it can be derived more carefully, and it turns out that $\gamma = \pi^2/3$, see Ashcroft and Mermin).

We can then derive the heat capacity

$$C = \partial E / \partial T = \gamma k_B g(E_F) k_B T V$$

which then using Eq. 4.11 we can rewrite as

$$C = \gamma \left(\frac{3Nk_B}{2} \right) \left(\frac{T}{T_F} \right)$$

The first term in brackets is just the classical result for the heat capacity of a gas, but the final factor T/T_F is tiny (0.01 or smaller!). This is the above promised linear T term in the specific heat of electrons, which is far smaller than one would get for a classical gas.

This Sommerfeld prediction for the electronic (linear T) contribution to the heat capacity of a metal is typically not far from being correct (The coefficient may be incorrect by factors of “order one”). A few metals, however, have specific heats that deviate from this prediction by as much as a factor of 10. Note that there are other measurements that indicate that these errors are associated with the electron mass being somehow changed in the metal. We will discover the reason for these deviations later when we study band theory (mainly in chapter 16).

Realizing now that the specific heat of the electron gas is reduced from that of the classical gas by a factor of $T/T_F \lesssim 0.01$, we can return to re-examine some of the above Drude calculations of thermal transport. We had above found (See Eq. 3.3-3.5) that Drude theory predicts a thermopower $S = \Pi/T = -c_v/(3e)$ that is too large by a factor of 100. Now it is clear that the reason for this error was that we used in this calculation (See Eq. 3.4) the specific heat per electron for a classical gas, which is too large by roughly $T_F/T \approx 100$. If we repeat the calculation using the proper specific heat, we will now get a prediction for thermopower which is reasonably close to what is actually measured in experiment for most metals.

We also used the specific heat per particle in the Drude calculation of the thermal conductivity $\kappa = \frac{1}{3} n c_v \langle v \rangle^2 \lambda$. In this case, the c_v that Drude used was too large by a factor of T_F/T , but on the other hand the value of $\langle v \rangle^2$ that he used was too small by roughly the same factor (Classically, one uses $mv^2/2 = k_B T$ whereas for the Sommerfeld model, one should use the Fermi velocity $mv_F^2/2 = k_B T_F$). Thus Drude’s prediction for thermal conductivity came out roughly correct (and thus the Wiedemann-Franz law correctly holds).

4.3 Magnetic Spin Susceptibility (Pauli Paramagnetism)¹¹

Another property we can examine about the free electron gas is its response to an externally applied magnetic field. There are several ways that the electrons can respond to the magnetic field. First, the electrons’ motion can be curved due to the Lorentz force. We have discussed this previously, and we will return to discuss it again in section 18.5 below¹². Secondly, the electron

¹¹Part VII of this book is entirely devoted to the subject of magnetism, so it might seem to be a bit out of place to discuss magnetism now. However since the calculation is an important result that hinges only on free electrons and Fermi statistics, it seems appropriate to me that it is discussed here. Most students will already be familiar with the necessary definitions of quantities such as magnetization and susceptibility so should not be confused by this. However, for those who disagree with this strategy or are completely confused by this section it is OK to skip over it and return after reading a bit of part VII.

¹²For a free electron gas, the contribution to the magnetic susceptibility from the orbital motion of the electron is known as Landau diamagnetism and takes the value $\chi_{Landau} = -(1/3)\chi_{Pauli}$. We will discuss diamagnetism

spins can flip over due to the applied magnetic field — this is the effect we will focus on. Roughly, the Hamiltonian (neglecting the Lorentz force of the magnetic field, see section 18.3 below for more detail) becomes¹³.

$$\mathcal{H} = \frac{\mathbf{p}^2}{2m} + g\mu_B \mathbf{B} \cdot \boldsymbol{\sigma}$$

where $g = 2$ is the g-factor of the electron¹⁴, \mathbf{B} is the magnetic field¹⁵ and $\boldsymbol{\sigma}$ is the spin of the electron which takes eigenvalues $\pm 1/2$. Here I have defined (and will use elsewhere) the useful version of the Bohr magneton

$$\mu_B = e\hbar/2m_e \approx .67(K/T)/k_B.$$

Thus in the magnetic field the energy of an electron with spin up or down (with up meaning it points the same way as the applied field, and $B = |\mathbf{B}|$)

$$\begin{aligned}\epsilon(\mathbf{k}, \uparrow) &= \frac{\hbar^2 |\mathbf{k}|^2}{2m} + \mu_B B \\ \epsilon(\mathbf{k}, \downarrow) &= \frac{\hbar^2 |\mathbf{k}|^2}{2m} - \mu_B B\end{aligned}$$

The spin magnetization of the system (moment per unit volume) in the direction of the applied magnetic field will then be

$$M = -\frac{1}{V} \frac{dE}{dB} = -([\# \text{ up spins}] - [\# \text{ down spins}]) \mu_B / V \quad (4.12)$$

So when the magnetic field is applied, it is lower energy for the spins to be pointing down, so more of them will point down. Thus a magnetization develops in the same direction as the applied magnetic field. This is known as *Pauli Paramagnetism*. Here *Paramagnetism* means that the magnetization is in the direction of the applied magnetic field. *Pauli Paramagnetism* refers in particular to the spin magnetization of the free electron gas. (We will discuss paramagnetism in more detail in chapter 18).

Let us now calculate the Pauli paramagnetism of the free electron gas at $T = 0$. With zero magnetic field applied, both the spin up and spin down states are filled up to the Fermi energy (i.e. to the Fermi wavevector). Near the Fermi level the density of states per unit volume for spin up electrons is $g(E_F)/2$ and similarly the density of states per unit volume for spin down electrons is $g(E_F)/2$. When B is applied, the spin ups will be more costly by an energy $\mu_B B$. Thus, (assuming that the chemical potential does not change) we will have $(g(E_F)/2)\mu_B B$ fewer spin up electrons

more in chapter 18 below. Unfortunately, calculating this diamagnetism is relatively tricky. (See Peierls' book for example). This effect is named after the famous Russian Nobel-Laureate Lev Landau, who kept a now famous ranking of how smart various physicist were — ranked on a logarithmic scale. Einstein was on top with a ranking of 0.5. Bose, Wigner, and Newton all received a ranking of 1. Schroedinger, Heisenberg, Bohr, and Dirac were ranked 2, and Landau modestly ranked himself a 2.5 but after winning the Nobel prize raised himself to 2. He said that anyone ranked below 4 was not worth talking to.

¹³The sign of the last term, the so called Zeeman coupling, may be a bit confusing. Recall that because the electron charge is negative, the electron dipole moment is actually opposite the direction of the electron spin (the current is rotating opposite the direction that the electron is spinning). Thus spins are lower energy when they are anti-aligned with the magnetic field! This is yet another annoyance caused by Benjamin Franklin who declared that the charge left on a glass rod when rubbed with silk is positive.

¹⁴It is a yet another constant source of grief that the letter “g” is used both for density of states and for g-factor of the electron. To avoid confusion we immediately set the g-factor to 2 and henceforth in this chapter g is reserved for density of states. Similar grief is that we now have to write \mathcal{H} for Hamiltonian because $H = B/\mu_0$ is frequently used for the magnetic field with μ_0 the permeability of free space.

¹⁵One should be careful to use the magnetic field seen by the actual electrons — this may be different from the magnetic field applied to the sample if the sample itself develops a magnetization.

per unit volume. Similarly, the spin downs will be less costly by the same amount, so we will have $(g(E_F)/2)\mu_B B$ more spin downs per unit volume. Note that the total number of electrons in the system did not change, so our assumption that the chemical potential did not change is correct. (Recall that chemical potential is always adjusted so it gives the right total number of electrons in the system). This process is depicted in Figure 4.3.

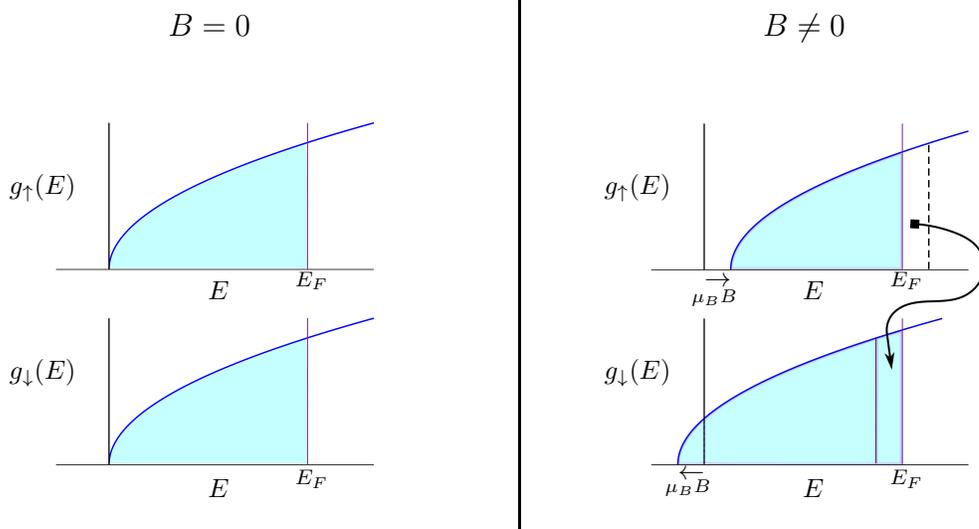


Figure 4.2: Left: Before the magnetic field is applied the density of states for spin up and spin down are the same $g_{\uparrow}(E) = g_{\downarrow}(E) = g(E)/2$. Note that these functions are proportional to $E^{1/2}$ (See Eq. 4.11) hence the shape of the curve, and the shaded region indicates the states that are filled. Right: When the magnetic field is applied, the states with up and down spin are shifted in energy by $+\mu_B B$ and $-\mu_B B$ respectively as shown. Hence up spins pushed above the Fermi energy can lower their energies by flipping over to become down spins. The number of spins that flip (the area of the approximately rectangular sliver) is roughly $g_{\uparrow}(E_F)\mu_B B$.

Using Eq. 4.12, given that we have moved $g(E_F)\mu_B B/2$ up spins to down spins, the magnetization (magnetic moment per unit volume) is given by

$$M = g(E_F)\mu_B^2 B$$

and hence the magnetic susceptibility $\chi = \partial M/\partial H$ is given (at $T = 0$ by)¹⁶

$$\chi_{Pauli} = \frac{dM}{dH} = \mu_0 \frac{dM}{dB} = \mu_0 \mu_B^2 g(E_F)$$

with μ_0 the permeability of free space. In fact this result is not far from correct for simple metals such as Li, Cu, or Na.

¹⁶See also the very closely related derivation given in section 22.1.2 below.

4.4 Why Drude Theory Works so Well

In retrospect we can understand a bit more about why Drude theory was so successful. As mentioned above, we now realize that because of Fermi statistics, treating electrons as a classical gas is incorrect – resulting in a huge overestimation of the heat capacity per particle, and in a huge underestimation of the typical velocity of particles. As described above, these two errors can sometimes cancel giving reasonable results nonetheless.

However, we can also ask why it is that Drude was successful in calculation of transport properties such as the conductivity and the Hall coefficient. In these calculations neither the velocity of the particle nor the specific heat enter. But still, the idea that a single particle will accelerate freely for some amount of time, then will scatter back to zero momentum seems like it must be wrong, since the state at zero momentum is always fully occupied. The transport equation (Eq. 3.1) that we solve

$$\frac{d\mathbf{p}}{dt} = \mathbf{F} - \frac{\mathbf{p}}{\tau} \quad (4.13)$$

in the Drude theory describes the motion of each particle. However, we can just as well use the same equation to describe the motion of the center of mass of the entire Fermi sea! On the left of Fig. 4.3 we have a picture of a Fermi sphere of radius \mathbf{k}_F . The typical electron has a very large velocity on the order of the Fermi velocity v_F , but the average of all of the (vector) velocities is zero. When an electric field is applied (in the \hat{y} direction as shown on the right of Fig. 4.3, so that the force is in the $-\hat{y}$ direction since the charge of the electron is $-e$) every electron in the system accelerates together in the $-\hat{y}$ direction, and the center of the Fermi sea shifts. The shifted Fermi sea has some nonzero average velocity, known as the *drift velocity* \mathbf{v}_{drift} . Since the kinetic energy of the shifted Fermi sea is higher than the energy of the Fermi sea with zero average velocity, the electrons will try to scatter back (with scattering rate $1/\tau$) to lower kinetic energy and shift the Fermi sea back to its original configuration with zero drift velocity. We can then view the Drude transport equation (Eq. 4.13) as describing the motion of the average velocity (momentum) of the entire Fermi sea.

One can think about how this scattering actually occurs in the Sommerfeld model. Here, most electrons have nowhere to scatter to, since all of the available \mathbf{k} states with lower energy (lower $|\mathbf{k}|$) are already filled. However, the few electrons near the Fermi surface in the thin crescent between the shifted and unshifted Fermi sea into the thin unfilled crescent on the other side of the unfilled Fermi sea to lower their energies (see Fig. 4.3). Although these scattering processes happen only to a very few of the electrons, the scattering events are extremely violent in that the change in momentum is exceedingly large (scattering all the way across the Fermi sea¹⁷).

4.5 Shortcomings of the Free Electron Model

Although the Sommerfeld (Free Electron) Model of a metal explains quite a bit about metals, it remains incomplete. Here are some items that are not well explained within Sommerfeld theory:

- Having discovered now that the typical velocity of electrons v_F is extremely large, and being able to measure the scattering time τ , we obtain a scattering length $\lambda = v_F\tau$ that may be 100 Angstroms or more. One might wonder, if there are atoms every few angstroms in a

¹⁷Actually, it may be that many small scatterings walking around the edge of these crescents make up this one effective scattering event

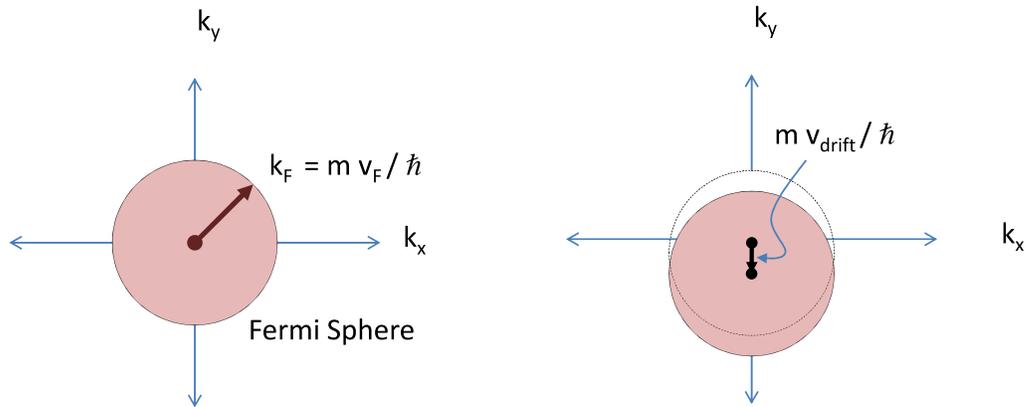


Figure 4.3: Drift Velocity and Fermi Velocity. The Drift momentum is the displacement of the entire Fermi sphere (which is generally very very small) whereas the Fermi momentum is the radius of the Fermi sphere, which can be very large. Drude theory makes sense if you think of it as a transport equation for the center of mass of the entire Fermi sphere – i.e., it describes the drift velocity. Scattering of electrons only occurs between the thin crescents that are the difference between the shifted and unshifted Fermi spheres

metal, why do the electrons not scatter from these atoms? (We will discuss this in chapter 14 below — the resolution is a result of Bloch’s theorem.)

- Many of our results depend on the number of electrons in a metal. In order to calculate this number we have always used the chemical valence of the atom. (For example, we assume one free electron per Li atom). However, in fact, except for Hydrogen, there are actually many electrons per atom. Why do core electrons not “count” for calculating the Fermi energy or velocity? What about insulators where there are no electrons free?
- We have still not resolved the question of why the Hall effect sometimes comes out with the incorrect sign, as if the charge carrier were positive rather than negative (the sign of charge of electrons.)
- In optical spectra of metals there are frequently many features (higher absorption at some frequencies, lower absorption at other frequencies). These features give metals their characteristic colors (for example, they make gold yellowish). The Sommerfeld model does not explain these features at all.
- The measured specific heat of electrons is much more correct than in Drude theory, but for

some metals is still off by factors as large as 10. Measurements of the mass of the electron in a metal also sometimes give answers that differ from the actual mass of the electron by similar factors.

- Magnetism: Some metals, such as Iron, are magnetic even without any applied external magnetic field. We will discuss magnetism in part VII below.
- Electron interaction: We have treated the electrons as noninteracting fermions. In fact, the typical energy of interaction for electrons, $e^2/(4\pi\epsilon_0 r)$ with r the typical distance between electrons) is huge, roughly the same scale as the Fermi energy. Yet we have ignored the Coulomb interaction between electrons completely. Understanding why this works is an extremely hard problem that was only understood starting in the late 1950s — again due to the brilliance of Lev Landau (See above footnote 12 in this chapter about Landau). The theory that explains this is frequently known as “Landau Fermi Liquid Theory”, but we will not study it in this course.

With the exception of the final two points (Magnetism and Electron interaction) all of these issues will be resolved once we study electronic *band structure* in chapters 10, 14 and particularly 16 below. In short, we are not taking seriously the periodic structure of atoms in materials.

4.6 Summary of (Sommerfeld) Free Electron Theory

- Treats properly the fact that electrons are Fermions.
- High density of electrons results in extremely high Fermi energy and Fermi velocity. Thermal and electric excitations are small redistributions of electrons around the Fermi surface.
- Compared to Drude theory, obtains electron velocity ~ 100 times larger, but heat capacity per electron ~ 100 times smaller. Leaves Wiedemann-Franz ratio roughly unchanged from Drude, but fixes problems in predications of thermal properties. Drude transport equations make sense if one considers velocities to be drift velocities, not individual electron velocities.
- Specific Heat and (Pauli) paramagnetic susceptibility can be calculated explicitly (know these derivations!) in good agreement with experiment.

References

For free electron (Sommerfeld) theory, good references are:

- Ashcroft and Mermin chapter 2–3.
- Singleton, section 1.5–1.6
- Rosenberg section 7.1–7.9
- Ibach and Luth section 6–6.5
- Kittel chapter 6
- Burns chapter 9B (excluding 9.14 and 9.16)

Part II

Putting Materials Together

Chapter 5

What Holds Solids Together: Chemical Bonding

In chapter 2 we found that the Debye model gave a reasonably good description of the specific heat of solids. However, we also found a number of shortcomings of the theory. These shortcomings basically stemmed from not taking seriously the fact that solids are actually made up of individual atoms assembled in a periodic structure.

Similarly in chapter 4 we found that the Sommerfeld model of metals described quite a bit about metals, but had a number of shortcomings as well — many of these were similarly due to not realizing that the solids are made up of individual atoms assembled in periodic structures.

As such, a large amount of this book will actually be devoted to understanding the effects of these individual atoms and their periodic arrangement on the electrons and on the vibrations of the solid. However, first it is worth backing up and asking ourselves why atoms stick together to form solids in the first place!

5.1 General Considerations about Bonding

To determine why atoms stick together to form solids, we are in some sense trying to describe the solution to a many particle Schroedinger¹ equation describing the many electrons and many nuclei in a solid. We can at least write down the equation

$$H\Psi = E\Psi$$

where Ψ is the wavefunction describing the positions and spin states of all the electrons and nuclei in the system. The terms in the Hamiltonian include a kinetic term (with inputs of the electron

¹Erwin Schroedinger was a fellow at Magdalen College Oxford from 1933 to 1938, but he was made to feel not very welcome there because he had a rather “unusual” personal life — he lived with both his wife, Anny, and with his mistress, Hilde, who, although married to another man, bore Schroedinger’s child, Ruth. After Oxford, Schroedinger was coaxed to live in Ireland with the understanding that this unusual arrangement would be fully tolerated. Surprisingly, all of the parties involved seemed fairly content until 1946 after Schroedinger fathered two more children with two different Irish women, whereupon Hilde decided to take Ruth back to Austria to live with her lawful husband. Anny, entirely unperturbed by this development and having her own lovers as well, remained Erwin’s close companion until his death.

and nucleon mass) as well as a Coulomb interaction term between all the electrons and nuclei.² While this type of description of chemical bonding is certainly true, it is also mostly useless. No one ever even tries to solve the Schroedinger equation for more than a few particles at a time. Trying to solve it for 10^{23} electrons simultaneously is completely absurd. One must try to extract useful information about the behavior from simplified models in order to obtain a qualitative understanding. (This is a great example of what I was ranting about in chapter 1 — reductionism does not work: saying that the Schroedinger equation is the whole solution is misguided). More sophisticated techniques try to turn these qualitative understandings into quantitative predictions.

In fact, what we are trying to do here is to try to understand a whole lot of chemistry from the point of view of a physicist. If you have had a good chemistry course, much of this chapter may sound familiar. However, here we will try to understand chemistry using our knowledge of quantum mechanics. Instead of learning empirical chemistry rules, we will look at simplified models that show roughly how these rules arise. However at the end of the day, we cannot trust our simplified models too much and we really should learn more chemistry to try to decide if yttrium really will form a carbonate salt or some similar question.

1	2										3	4	5	6	7	0	
																He	
Li	Be										B	C	N	O	F	Ne	
Na	Mg										Al	Si	P	S	Cl	Ar	
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
Cs	Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
Fr	Ra	Ac	Rf	Db	Sg	Bh	Hs	Mt	Ds	Rg							

Legend:

- Alkali metals (blue)
- Transition metals (yellow)
- Halogens (green)
- Noble gases (red)

Figure 5.1: The periodic table of the elements.

From a chemist’s point of view one frequently thinks about different *types* of chemical bonds depending on the types of atoms involved, and in particular, depending on the atom’s position on the periodic table (and in particular, on the atom’s *electronegativity* — which is its tendency to attract electrons). Below we will discuss Ionic Bonds, Covalent Bonds, van der Waals (fluctuating dipole, or molecular) bonds, Metallic Bonds, and Hydrogen Bonds. Of course, they are all different aspects of the Schroedinger equation, and any given material may exhibit aspects of several of these types of bonding. Nonetheless, qualitatively it is quite useful to discuss these different types of bonds to give us intuition about how chemical bonding can occur. A brief description of the many types of bonding and their properties is shown in table 5.1. Note that this table should be considered just as rules-of-thumb, as many materials have properties intermediate between the categories listed.

²To have a fully functioning “Theory of Everything” as far as all of chemistry, biology, and most of everything that matters to us (besides the sun and atomic energy) is concerned, one needs only Coulomb interaction plus the Kinetic term in the Hamiltonian, plus spin-orbit (relativistic effects) for some of the heavy atoms.

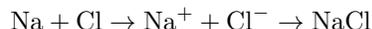
Type of Bonding	Description	Typical of which compounds	Typical Properties
Ionic	Electron is transferred from one atom to another, and the resulting ions attract each other	Binary compounds made of constituents with very different electronegativity: Ex, group I-VII compounds such as NaCl.	<ul style="list-style-type: none"> • Hard, Very Brittle • High Melting Temperature • Electrical Insulator • Water Soluble
Covalent	Electron is shared equally between two atoms forming a bond. Energy lowered by delocalization of wavefunction	Compounds made of constituents with similar electronegativities (ex, III-V compounds such as GaAs), or solids made of one element only such as diamond (C)	<ul style="list-style-type: none"> • Very Hard (Brittle) • High Melting Temperature • Electrical Insulators or Semiconductors
Metallic Bonds	Electrons delocalized throughout the solid forming a glue between positive ions.	Metals. Left and Middle of Periodic Table.	<ul style="list-style-type: none"> • Ductile, Maleable (due to non-directional nature of bond. Can be hardened by preventing dislocation motion with impurities) • Lower Melting Temperature • Good electrical and thermal conductors.
Molecular (van der Waals or Fluctuating Dipole)	No transfer of electrons. Dipole moments on constituents align to cause attraction. Bonding strength increases with size of molecule or polarity of constituent.	Noble Gas Solids, Solids made of Non-Polar (or slightly polar) Molecules Binding to Each Other (Wax)	<ul style="list-style-type: none"> • Soft, Weak • Low Melting Temperature • Electrical Insulators
Hydrogen	Involves Hydrogen ion bound to one atom but still attracted to another. Special case because H is so small.	Important in organic and biological materials	<ul style="list-style-type: none"> • Weak Bond (stronger than VdW though) • Important for maintaining shape of DNA and proteins

Table 5.1: Types of Bonds in Solids. This table should be thought of as providing rough rules. Many materials show characteristics intermediate between two (or more!) classes.

In this section we will try to be a bit more quantitative about how some of these types of bonding come about. Remember, underneath it is all the Schroedinger equation and the Coulomb interaction between electrons and nuclei that is holding materials together!

5.2 Ionic Bonds

The general idea of an ionic bond is that for certain compounds (for example, binary compounds, such as NaCl, made of one element in group I and one element in group VII), it is energetically favorable for an electron to be physically transferred from one atom to the other, leaving two oppositely charged ions which then attract each other. One writes a chemical “reaction” of the form



To find out if such a reaction happens, one must look at the energetics associated with the transfer of the electron.

At least in principle it is not *too* hard to imagine solving the Schroedinger equation³ for a single atom and determining the energy of the neutral atom, of the positive ion, and of the negative ion or actually measuring these energies for individual atoms with some sort of spectroscopy. We define:

Ionization Energy = Energy required to remove one electron
from a neutral atom to create a positive ion

Electron Affinity = Energy gain for creating a negative ion
from a neutral atom by adding an electron

To be precise, in both cases we are comparing the energy of having an electron either at position infinity, or on the atom. Further, if we are removing or adding only a single electron, then these are called *first* Ionization energies and *first* electron affinities respectively (one can similarly define energies for removing or adding two electrons which would be called *second*). Finally we note that chemists typically work with systems at fixed (room) temperature and (atmospheric) pressure, in which case they are likely to be more concerned with Gibbs free energies, rather than pure energies. We will always assume that one is using the appropriate free energy for the experiment in question (and we will be sloppy and always call an energy E).

Ionization energy is smallest on the left (group I and II) of the periodic table and largest on the right (group VII, and VIII). To a lesser extent the ionization energy also tends to decrease towards the bottom of the periodic table. Similarly electron affinity is also largest on the right and top of the periodic table (not including the group VIII noble gases which roughly do not attract electrons measurably at all).

The total energy change from transferring an electron from atom A to atom B is

$$\Delta E_{A+B \rightarrow A^+ + B^-} = (\text{Ionization Energy})_A - (\text{Electron Affinity})_B$$

³As emphasized in chapter 1 even the world’s largest computers cannot solve the Schroedinger equation for a system of more than a few electrons. Nobel prizes (in chemistry) were awarded to Walter Kohn and John Pople for developing computational methods that can obtain highly accurate approximations. These approaches have formed much of the basis of modern quantum chemistry.

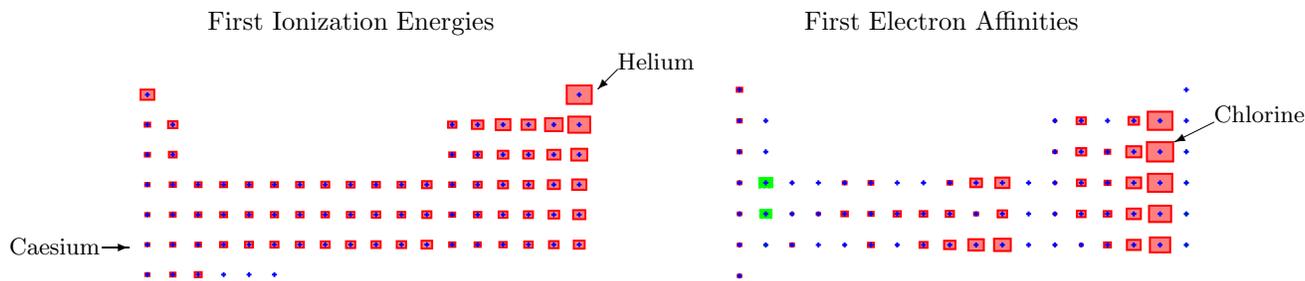


Figure 5.2: Pictorial Tables of First Ionization Energies (left) and First Electron Affinities (right). The word "First" here means that we are measuring the energy to lose or gain a first electron starting with a neutral atom. The linear size of each box represents the magnitude of the energies (scales on the two plots differ). For reference the largest ionization energy is helium, at roughly 24.58 eV per atom, the lowest is caesium at 3.89 eV. The largest electron affinity is chlorine which gains 3.62 eV when binding to an additional electron. The few light green colored boxes are atoms that have *negative* electron affinities.

Note carefully the sign. The ionization energy is a positive energy that must be put in, the electron affinity is an energy that comes out.

However this ΔE is the energy to transfer an electron between two atoms very far apart. In addition, there is also⁴

$$\text{Cohesive Energy} = \text{Energy gain from } A^+ + B^- \rightarrow AB$$

This cohesive energy is mostly a classical effect of the Coulomb interaction between the ions as one lets the ions come close together.⁵ The total energy gain for forming a molecule from the two individual atoms is thus given by

$$\Delta E_{A+B \rightarrow AB} = (\text{Ionization Energy})_A - (\text{Electron Affinity})_B - \text{Cohesive Energy of A-B}$$

One obtains an ionic bond if the total ΔE for this process is less than zero.

In order to determine whether an electron is likely to be transferred between one atom and another, it is convenient to use the so-called *electronegativity*, which roughly describes how much an atom "wants" electrons, or how much an atom attracts electrons to itself. While there are

⁴The term "Cohesive Energy" can be ambiguous since sometimes people use it to mean the energy to put two ions together into a compound, and other times they mean it to be the energy to put two neutral atoms together! Here we mean the former.

⁵One can write a simple classical equation for a total cohesive energy for a solid

$$E_{\text{cohesive}} = - \sum_{i < j} \frac{Q_i Q_j}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{r}_j|}$$

where Q_i is the charge on the i^{th} ion, and \mathbf{r}_i is its position. This sum is sometimes known as the *Madelung Energy*. It might look like one could make the cohesive energy infinitely large by letting two ions come to the same position! However, when atoms approach each other within roughly an atomic radius there is an additional strong repulsion associated with the Pauli exclusion principle that no two electrons may occupy the same orbital. One thus needs a more quantum mechanical treatment to determine, ab-initio, how close two oppositely charged ions will come to each other.

various definitions of electronegativity that are used, a simple and useful definition is known as the Mulliken Electronegativity^{6,7}

$$\text{(Mulliken) Electronegativity} = \frac{(\text{Electron Affinity}) + (\text{Ionization Energy})}{2}$$

The electronegativity is extremely large for elements in the upper right of the periodic table (not including the noble gases).

In bonding, the electron is always transferred from the atom of lower electronegativity to higher electronegativity. The greater the difference in electronegativities between two atoms the more completely the electron is transferred from one atom to the other. If the difference in electronegativities is small, then the electron is only partially transferred from one atom to the other. We will see below that one can have covalent bonding even between two identical atoms where there is no difference in electronegativities, and therefore no net transfer of electrons. Before leaving the topic of ionic bonds, it is worth discussing some of the typical physics of ionic solids. First of all, the materials are typically hard, as the Coulomb interaction between oppositely charged ions is strong. However, since water is extremely polar, it can dissolve an ionic solid. This happens (See Fig 5.3) by arranging the water molecules such that the negative side of the molecule is close to the positive ions and the positive side of the molecule is close to the negative ions.

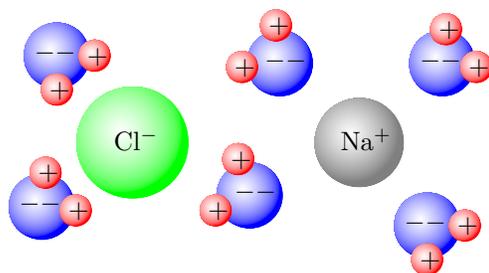


Figure 5.3: Salt, NaCl, dissolved in water. Ionic compounds typically dissolve easily in water since the polar water molecules can screen the highly charged, but otherwise stable, ions.

⁶This electronegativity can be thought as approximately the negative of the chemical potential via

$$\frac{1}{2}(E_{\text{affinity}} + E_{\text{ion}}) = \frac{1}{2}([E_N - E_{N+1}] + [E_{N-1} - E_N]) = \frac{E_{N-1} - E_{N+1}}{2} \approx -\frac{\partial E}{\partial N} \approx -\mu.$$

See however the comments in section 4.1 on defining chemical potential for systems with discrete energy levels and discrete number of electrons.

⁷Both Robert Mulliken and Linus Pauling won Nobel Prizes in Chemistry for their work understanding chemical bonding including the concept of electronegativity. Pauling won a second Nobel prize, in Peace, for his work towards banning nuclear weapons testing. (Only four people have ever won two Nobels: Marie Curie, Linus Pauling, John Bardeen, and Fredrick Sanger. We should all know these names!). Pauling was criticized later in his life for promoting high doses of vitamin C to prevent cancer and other ailments, sometimes apparently despite scientific evidence to the contrary.

5.3 Covalent Bond

Roughly, a covalent bond is a bond where electrons are shared equally between two atoms. There are several pictures that can be used to describe the covalent bond.

5.3.1 Particle in a Box Picture

Let us model a hydrogen atom as a box of size L for an electron (for simplicity, let us think about a one dimensional system). The energy of a single electron in a box is (I hope this looks familiar!)

$$E = \frac{\hbar^2 \pi^2}{2mL^2}$$

Now suppose two such atoms come close together. An electron that is shared between the two atoms can now be delocalized over the positions of both atoms, thus it is in a box of size $2L$ and has lower energy

$$E = \frac{\hbar^2 \pi^2}{2m(2L)^2}$$

This reduction in energy that occurs by delocalizing the electron is the driving force for forming the chemical bond. The new ground state orbital is known as a *bonding* orbital.

If each atom starts with a single electron (i.e., it is a hydrogen atom) then when the two atoms come together to form a lower energy (bonding) orbital, then *both* electrons can go into this same ground state orbital since they can take opposite spin states. Of course the reduction in energy of the two electrons must compete against the Coulomb repulsion between the two nuclei, and the Coulomb repulsion of the two electrons with each other, which is a much more complicated calculation.

Now suppose we had started with two helium atoms, where each atom has two electrons, then when the two atoms come together there is not enough room in the single ground state wavefunction. In this case, two of the four electrons must occupy the first excited orbital — which in this case turns out to be exactly the same electronic energy as the original ground state orbital of the original atoms — since no energy is gained by these electrons when the two atoms come together these are known as *antibonding* orbitals. (In fact it requires energy to push the two atoms together if one includes Coulomb repulsions between the nuclei)

5.3.2 Molecular Orbital or Tight Binding Theory

In this section we make slightly more quantitative some of the idea of the previous section. Let us write a Hamiltonian for two Hydrogen atoms. Since the nuclei are heavy compared to the electrons, we will fix the nuclear positions and solve the Schroedinger equation for the electrons as a function of the distance between the nuclei. This fixing of the position of nuclei is known as a

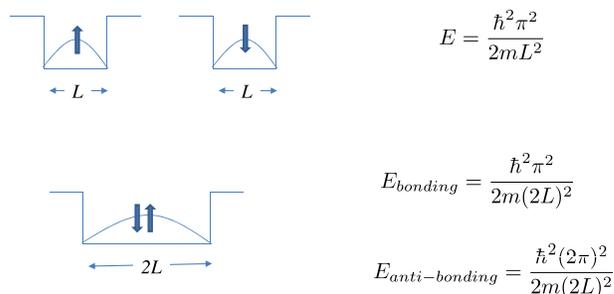


Figure 5.4: Particle in a box picture of covalent bonding. Two separated hydrogen atoms are like two different boxes each with one electron in the lowest eigenstate. When the two boxes are pushed together, one obtains a larger box – thereby lowering the energy of the lowest eigenstate – which is known as the *bonding* orbital. The two electrons can take opposite spin states and can thereby both fit in the bonding orbital. The first excited state is known as the *antibonding* orbital

“Born-Oppenheimer” approximation^{8,9}. We hope to calculate the eigenenergies of the system as a function of the distance between the positively charged nuclei.

For simplicity, let us consider a single electron and two identical positive nuclei. We write the Hamiltonian as

$$H = K + V_1 + V_2$$

with

$$K = \frac{\mathbf{p}^2}{2m}$$

being the kinetic energy of the electron and

$$V_i = \frac{e^2}{4\pi\epsilon_0|\mathbf{r} - \mathbf{R}_i|}$$

is the Coulomb interaction energy between the electron position \mathbf{r} and the position of nuclei \mathbf{R}_i .

Generally this type of Schroedinger equation is hard to solve exactly. (In fact it can be solved exactly in this case, but it is not particularly enlightening to do so). Instead, we will attempt a variational solution. Let us write a trial wavefunction as

$$|\psi\rangle = \phi_1|1\rangle + \phi_2|2\rangle \quad (5.1)$$

⁸Max Born (also the same guy from Born-Von Karmen boundary conditions) was one of the founders of quantum physics, winning a Nobel Prize in 1954. His daughter, and biographer, Irene, married into the Newton-John family, and had a daughter named Olivia, who became a pop icon and film star in the 1970s. Her most famous role was in the movie of *Grease* playing Sandra-Dee opposite John Travolta. When I was a kid, she was every teenage guy’s dream-girl (her, or Farrah Fawcett).

⁹J. Robert Oppenheimer later became the head scientific manager of the American atomic bomb project during the second world war. After this giant scientific and military triumph, he pushed for control of nuclear weapons leading to his being accused of being a communist sympathizer during the “Red” scares of the 1950s and he ended up having his security clearance revoked.

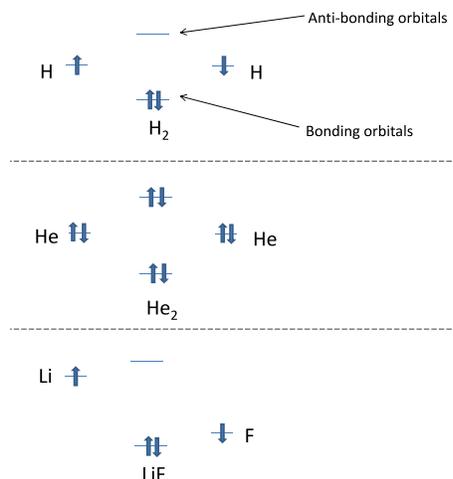


Figure 5.5: Molecular Orbital Picture of Bonding. In this type of picture, on the far left and far right are the orbital energies of the individual atoms well separated from each other. In the middle are the orbital energies when the atoms come together to form a molecule. Top: Two hydrogen atoms come together to form a H₂ molecule. As mentioned above in the particle-in-a-box picture, the lowest energy eigenstate is reduced in energy when the atoms come together and both electrons go into this bonding orbital. Middle: In the case of helium, since there are two electrons per atom, the bonding orbitals are filled, and the antibonding orbitals must be filled as well. The total energy is not reduced by the two Helium atoms coming together (thus helium does not form He₂). Bottom: In the case of LiF, the energies of the lithium and the fluorine orbitals are different. As a result, the bonding orbital is mostly composed of the orbital on the Li atom – meaning that the bonding electrons are mostly transferred from Li to F — forming a more ionic bond.

where ϕ_i are complex coefficients, and the kets $|1\rangle$ and $|2\rangle$ are known as “atomic orbitals” or “tight binding” orbitals¹⁰. The form of Eq. 5.1 is frequently known as a “linear combination of atomic orbitals” or LCAO¹¹. The orbitals which we use here can be taken as the ground state solution of the Schrodinger equation when there is only one nucleus present. I.e.

$$\begin{aligned} (K + V_1)|1\rangle &= \epsilon_0|1\rangle \\ (K + V_2)|2\rangle &= \epsilon_0|2\rangle \end{aligned} \quad (5.2)$$

where ϵ_0 is the ground state energy of the single atom¹². I.e., $|1\rangle$ is a ground state orbital on

¹⁰The term “tight binding” is from the idea that an atomic orbital is tightly bound to its nucleus.

¹¹The LCAO approach can be improved systematically by using more orbitals and more variational coefficients — which then can be optimized with the help of a computer. This general idea formed the basis of the quantum chemistry work of John Pople. See footnote 3 above in this section.

¹²Here ϵ_0 is *not* a dielectric constant or the permittivity of free space, but rather the energy of an electron in an orbital (At some point we just run out of new symbols to use for new quantities!)

nucleus 1 and $|2\rangle$ is a ground state orbital on nucleus 2.

For simplicity, we will now make a rough approximation that $|1\rangle$ and $|2\rangle$ are orthogonal so we can then choose a normalization such that

$$\langle i|j\rangle = \delta_{ij} \quad (5.3)$$

When the two nuclei get very close together, this orthogonality is clearly no longer even close to correct. We then have to decide: either we keep our definition of the atomic orbitals being the solution to the Schroedinger equation for a single nucleus, but we give up on the two atomic orbitals being orthogonal; or we can give up on the orbitals being solutions to the Schroedinger equation for a single nucleus, but we keep orthonormality. It is a good exercise to consider what happens when we give up orthonormality, but fortunately most of what we learn does not depend too much on whether the orbitals are orthogonal or not, so for simplicity we will assume orthonormal orbitals.

An effective Schroedinger equation can be written down for our variational wavefunction which (unsurprisingly) takes the form of an eigenvalue problem¹³

$$\sum_j H_{ij}\phi_j = E\phi_i$$

where

$$H_{ij} = \langle i|H|j\rangle$$

is a two by two matrix in this case. (The equation generalizes in the obvious way to the case where there are more than 2 orbitals).

Recalling our definition of $|1\rangle$ as being the ground state energy of $K + V_1$, we can write¹⁴

$$H_{11} = \langle 1|H|1\rangle = \langle 1|K + V_1|1\rangle + \langle 1|V_2|1\rangle = \epsilon_0 + V_{cross} \quad (5.4)$$

$$H_{22} = \langle 2|H|2\rangle = \langle 2|K + V_2|2\rangle + \langle 2|V_1|2\rangle = \epsilon_0 + V_{cross} \quad (5.5)$$

$$H_{12} = \langle 1|H|2\rangle = \langle 1|K + V_2|2\rangle + \langle 1|V_1|2\rangle = 0 - t \quad (5.6)$$

$$H_{21} = \langle 2|H|1\rangle = \langle 2|K + V_2|1\rangle + \langle 2|V_1|1\rangle = 0 - t^* \quad (5.7)$$

In the first two lines

$$V_{cross} = \langle 1|V_2|1\rangle = \langle 2|V_1|2\rangle$$

is the Coulomb potential felt by orbital $|1\rangle$ due to nucleus 2, or equivalently the Coulomb potential felt by orbital $|2\rangle$ due to nucleus 1. In the second two lines (Eqs. 5.6 and 5.7) we have also defined the so-called *hopping* term^{15,16}

$$t = -\langle 1|V_2|2\rangle = -\langle 1|V_1|2\rangle$$

¹³To derive this eigenvalue equation we start with an expression for the energy

$$E = \frac{\langle \psi|H|\psi\rangle}{\langle \psi|\psi\rangle}$$

then with ψ written in the variational form of Eq. 5.1, we minimize the energy by setting $\partial E/\partial\phi_i = \partial E/\partial\phi_i^* = 0$.

¹⁴In atomic physics courses, the quantities V_{cross} and t are often called a *direct* and *exchange* terms and are sometimes denoted \mathcal{J} and \mathcal{K} . We avoid this terminology because the same words are almost always used to describe 2-electron interactions in condensed matter.

¹⁵The minus sign is a convention for the definition of t . For many cases of interest, this definition makes t positive, although it can actually have either sign depending on the structure of the orbitals in question and the details of the potential.

¹⁶The second equality here can be obtained by rewriting $H_{12} = \langle 1|K + V_1|2\rangle + \langle 1|V_2|2\rangle$.

The reason for the name “hopping” will become clear below. Note that in the second two lines (Eqs. 5.6 and 5.7) the first term vanishes because of orthogonality of $|1\rangle$ and $|2\rangle$. Thus our Schroedinger equation is reduced to a two by two matrix equation of the form

$$\begin{pmatrix} \epsilon_0 + V_{cross} & -t \\ -t^* & \epsilon_0 + V_{cross} \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = E \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \quad (5.8)$$

The interpretation of this equation is roughly that orbitals $|1\rangle$ and $|2\rangle$ both have energies ϵ_0 which is shifted by V_{cross} due to the presence of the other nucleus. In addition the electron can “hop” from one orbital to the other by the off-diagonal t term. To understand this interpretation more fully, we realize that in the *time dependent* Schroedinger equation, if the matrix were diagonal a wavefunction that started completely in orbital $|1\rangle$ would stay on that orbital for all time. However, with the off-diagonal term, the time dependent wavefunction can oscillate between the two orbitals.

Diagonalizing this two-by-two matrix we obtain eigenenergies

$$E_{\pm} = \epsilon_0 + V_{cross} \pm |t|$$

the lower energy orbital is the bonding orbital whereas the higher energy orbital is the anti-bonding. The corresponding wavefunctions are then

$$\psi_{bonding} = \frac{1}{\sqrt{2}}(\phi_1 \pm \phi_2) \quad (5.9)$$

$$\psi_{anti-bonding} = \frac{1}{\sqrt{2}}(\phi_1 \mp \phi_2) \quad (5.10)$$

I.e., these are the symmetric and antisymmetric superposition of orbitals. The signs \pm and \mp depend on the sign of t , where the lower energy one is always called the bonding orbital and the higher energy one is called antibonding. To be precise $t > 0$ makes $(\phi_1 + \phi_2)/\sqrt{2}$ the lower energy bonding orbital. Roughly one can think of these two wavefunctions as being the lowest two “particle-in-a-box” orbitals — the lowest energy wavefunction does not change sign as a function of position, whereas the first excited state changes sign once, i.e., it has a single node (for the case of $t > 0$ the analogy is precise).

It is worth briefly considering what happens if the two nuclei being bonded together are not identical. In this case the energy ϵ_0 for an electron to sit on orbital 1 would be different from that of orbital 2. (See bottom of Fig. 5.5) The matrix equation 5.8 would no longer have equal entries along the diagonal, and the magnitude of ϕ_1 and ϕ_2 would no longer be equal in the ground state as they are in Eq. 5.9. Instead, the lower energy orbital would be more greatly filled in the ground state. As the energies of the two orbitals become increasingly different, the electron is more completely transferred entirely onto the lower energy orbital, essentially reducing to an ionic bond.

Aside: In section 22.4 below, we will consider a more general tight binding model with more than one electron in the system and with Coulomb interactions between electrons as well. That calculation is more complicated, but shows very similar results. That calculation is also much more advanced, but might be fun to read for the adventurous.

Note again that V_{cross} is the energy that the electron on orbital 1 feels from nucleus 2. However, we have not included the fact that the two nuclei also interact, and to a first approximation, this Coulomb repulsion between the two nuclei will cancel¹⁷ the attractive energy between the

¹⁷If you think of a positively charged nucleus and a negatively charged electron surrounding the nucleus, from far outside of that electron’s orbital radius the atom looks neutral. Thus a second nucleus will neither be attracted nor repelled from the atom so long as it remains outside of the electron cloud of the atom.

nucleus and the electron on the opposite orbital. Thus, including this energy we will obtain

$$\tilde{E}_{\pm} \approx \epsilon_0 \pm |t|$$

As the nuclei get closer together, the hopping term $|t|$ increases, giving an energy level diagram as shown in Fig. 5.3.2. This picture is obviously unrealistic, as it suggests that two atoms should bind together at zero distance between the nuclei. The problem here is that our assumptions and approximations begin to break down as the nuclei get closer together (for example, our orbitals are no longer orthogonal, V_{cross} does not exactly cancel the Coulomb energy between nuclei, etc.).

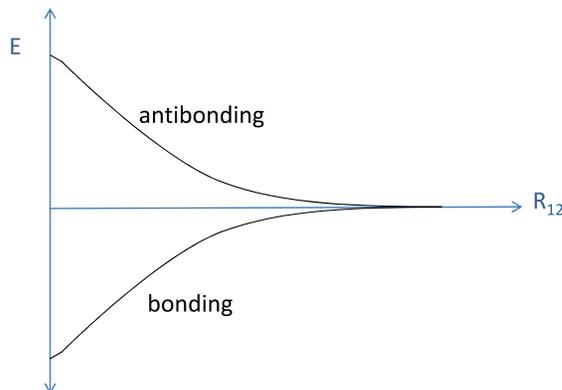


Figure 5.6: Model Tight Binding Energy Levels as a Function of Distance Between the Nuclei of the Atoms.

A more realistic energy level diagram for the bonding and antibonding states is given in Fig. 5.7. Note that the energy diverges as the nuclei get pushed together (this is from the Coulomb repulsion between nuclei). As such there is a minimum energy of the system when the nuclei are at some nonzero distance apart from each other, which then becomes the ground state distance of the nuclei in the resulting molecule.

Aside: In Fig. 5.7 there is a minimum of the bonding energy when the nuclei are some particular distance apart. This optimal distance will be the distance of the bond between two atoms. However, at finite temperature, the distance will fluctuate around this minimum (think of a particle in a potential well at finite temperature). Since the potential well is steeper on one side than on the other, at finite temperature, the “particle” in this well will be able to fluctuate to larger distances a bit more than it is able to fluctuate to smaller distances. As a result, the average bond distance will increase at finite temperature. This thermal expansion will be explored again in the next chapter.

Covalently bonded materials tend to be strong and tend to be electrical semiconductors or insulators (since electrons are tied up in the local bonds). The directionality of the orbitals makes these materials retain their shape well (non-ductile) so they are brittle. They do not dissolve in polar solvents such as water in the same way that ionic materials do.

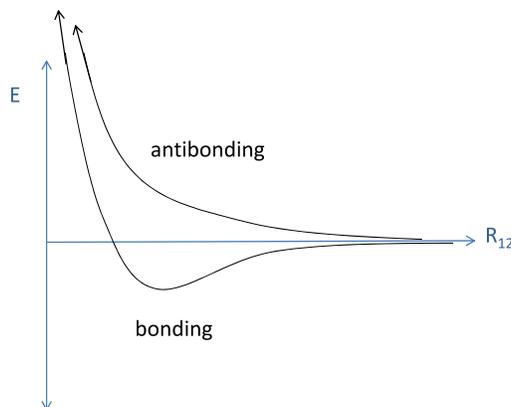


Figure 5.7: More Realistic Energy Levels as a Function of Distance Between the Nuclei of the Atoms.

5.4 Van der Waals, Fluctuating Dipole Forces, or Molecular Bonding

When two atoms (or two molecules) are very far apart from each other, there remains an attraction between them due to what is known as van der Waals¹⁸ forces, sometimes known as fluctuating dipole forces, or molecular bonding. In short, both atoms have a dipole moment, which may be zero on average, but can fluctuate momentarily due to quantum mechanics. If the first atom obtains a momentary dipole moment, the second atom can polarize — also obtaining a dipole moment to lower its energy. As a result, the two atoms (momentarily dipoles) will attract each other.

This type of bonding between atoms is very typical of inert atoms (such as noble gases: He, Ne, Kr, Ar, Xe) whose electrons do not participate in covalent bonds or ionic bonds. It is also typical of bonding between inert¹⁹ molecules such as nitrogen molecules N_2 where there is no possibility for the electrons in this molecule to form covalent or ionic bonds between molecules. This bonding is weak compared to covalent or ionic bonds, but it is also long ranged in comparison since the electrons do not need to hop between atoms.

To be more quantitative, let us consider an electron orbiting a nucleus (say, a proton). If the electron is at a fixed position, there is a dipole moment $\mathbf{p} = e\mathbf{r}$ where \mathbf{r} is the vector from the electron to the proton. With the electron “orbiting” (i.e. in an eigenstate), the average dipole moment is zero. However, if an electric field is applied to the atom, the atom will develop a polarization (i.e., it will be more likely for the electron to be found on one side of the nucleus than on the other). We write

$$\mathbf{p} = \chi\mathbf{E}$$

¹⁸J. D. van der Waals was awarded the Nobel prize in Physics in 1910 for his work on the structure of Liquids and Gases. You may remember the van der Waals equation of state from your thermodynamics course last year. There is a crater named after him on the far side of the moon.

¹⁹Whereas the noble gases are inert because they have filled atomic orbital shells, the nitrogen molecule is inert essentially because it has a filled shell of *molecular* orbitals — all of the bonding orbitals are filled, and there is a large energy gap to any anti-bonding orbitals.

where χ is known as the polarizability (also known as electric susceptibility). This polarizability can be calculated, for, say a hydrogen atom explicitly²⁰. At any rate, it is some positive quantity.

Now, let us suppose we have two such atoms, separated a distance r in the \hat{x} direction. Suppose one atom momentarily has a dipole moment \mathbf{p}_1 (for definiteness, suppose this dipole moment is in the \hat{z} direction). Then the second atom will feel an electric field

$$E = \frac{p_1}{4\pi\epsilon_0 r^3}$$

in the negative \hat{z} direction. The second atom then, due to its polarizability, develops a dipole moment $p_2 = \chi E$ which in turn is attracted to the first atom. The potential energy between these two dipoles is

$$U = \frac{-|p_1||p_2|}{4\pi\epsilon_0 r^3} = \frac{-p_1\chi E}{(4\pi\epsilon_0)r^3} = \frac{-|p_1|^2\chi}{(4\pi\epsilon_0 r^3)^2} \quad (5.11)$$

corresponding to a force $-dU/dr$ which is attractive and proportional to $1/r^7$.

You can check that independent of the direction of the original dipole moment, the force is always attractive and proportional to $1/r^7$. Although there will be a (nonnegative) prefactor which depends on the angle between the dipole moment \mathbf{p}_1 and \mathbf{x} the direction between the two atoms.

Note. This argument appears to depend on the fact that the dipole moment \mathbf{p}_1 of the first atom is nonzero. On average the atom's dipole moment will be zero. However in Eq. 5.11 in fact what enters is $|\mathbf{p}_1|^2$ which has a nonzero expectation value. (In fact this is precisely the calculation that $\langle x \rangle$ for an electron in a hydrogen atom is zero, but $\langle x^2 \rangle$ is nonzero).

While these fluctuating dipolar forces are generally weak, they are the only forces that occur when electrons cannot be shared or transferred between atoms — either in the case where the electrons are not chemically active or when the atoms are far apart. However, when considering the van der Waals forces of many atoms put together, the total forces can be quite strong. A well known example of a van der Waals force is the force that allows lizards, such as Geckos to climb up walls. They have hair on their feet that makes very close contact with the atoms of the wall, and they can climb up the walls mostly due to van der Waals forces!

5.5 Metallic Bonding

It is sometimes hard to distinguish metallic bonding from covalent bonding. Roughly, however, one defines a metallic bond to be the bonding that occurs in metal. These bonds are similar to covalent bonds in the sense that electrons are shared between atoms, but in this case the electrons become delocalized throughout the crystal (we will discuss how this occurs in section 10.2 below). We should think of the delocalized free electrons as providing the glue that holds together the positive ions that they have left behind.

Since the electrons are completely delocalized, the bonds in metals tend not to be directional. Metals are thus often ductile and malleable. Since the electrons are free, metals are good conductors of electricity as well as of heat.

²⁰This is a good exercise in quantum mechanics. See, for example, Eugen Merzbacher's book on quantum mechanics.

5.6 Hydrogen bonds

The hydrogen atom is extremely special due to its very small size. As a result, the bonds formed with hydrogen atoms are qualitatively different from other bonds. When the hydrogen atom forms a covalent or ionic bond with a larger atom, being small, the hydrogen nucleus (a proton) simply sits on the surface of its partner. This then makes the molecule (hydrogen and its partner) into a dipole. These dipoles can then attract charges, or other dipoles, as usual.

What is special about hydrogen is that when it forms a bond, and its electron is attracted away from the proton onto (or partially onto) its partner, the unbonded side of the the proton left behind is a naked positive charge – unscreened by any electrons in core orbitals. As a result, this positive charge is particularly effective in being attracted to other clouds of electrons.

A very good example of the hydrogen bond is water, H_2O . Each oxygen atom is bound to two hydrogens (however because of the atomic orbital structure, these atoms are not collinear). The hydrogens, with their positive charge remain attracted to oxygens of other water molecules. In ice, these attractions are strong enough to form a weak, but stable bond between water molecules, thus forming a crystal. Sometimes one can think of the the hydrogen atom as forming “half” a bond with two oxygen atoms, thus holding the two oxygen atoms together.

Hydrogen bonding is extremely important in biological molecules where, for example, hydrogen bonds hold together strands of DNA.

5.7 Summary of Bonding (Pictorial)

See also the table 5.1 for a summary of bonding types.

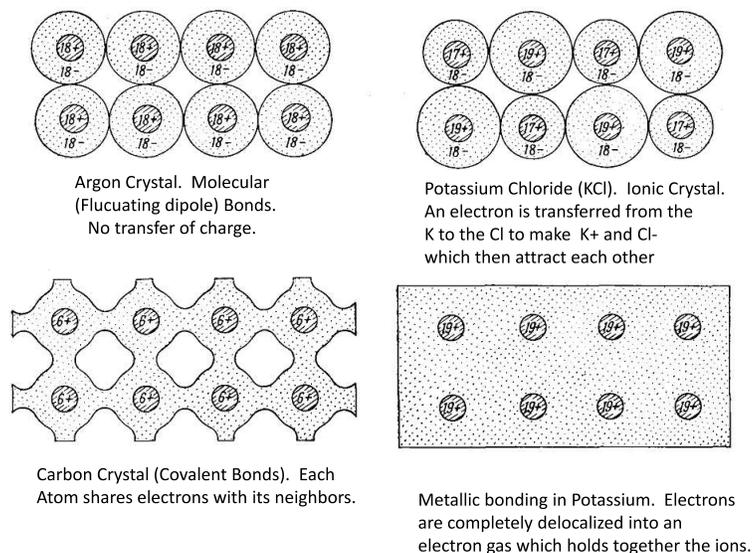


Figure 5.8: Cartoons of Bonding Types

References on Chemical Bonding

- Rosenberg, section 1.11–1.19
- Ibach and Luth, chapter 1
- Hook and Hall, section 1.6
- Kittel, chapter 3 up to elastic strain
- Ashcroft and Mermin, chapters 19–20
- Burns, section 6.2–6.6 and also chapters 7 and 8

Probably Ashcroft and Mermin as well as Burns chapters 7, and 8 are too much information.

Chapter 6

Types of Matter

Once we understand how it is that atoms bond together, we can examine what types of matter can be formed. An obvious thing that can happen is that atoms can bond together to form regular crystals. A crystal is made of small units reproduced many times and built into a regular array. The macroscopic morphology of a crystal can reflect its underlying structure (See Fig. 6) We will spend much of the remainder of this book studying crystals.

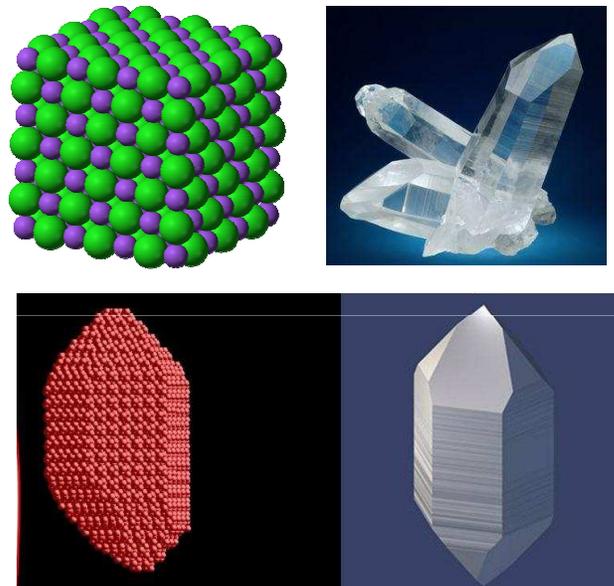


Figure 6.1: Crystals: Top left: Small units (One green one blue) reproduced periodically to form a crystal. Top right: A crystal of quartz (SiO_2). Bottom: The macroscopic morphology of a crystal reflects its underlying structure.

It is also possible that atoms will bind together to form molecules, and the molecules will stick together via weak Van der Waals bonds to form so-called *molecular crystals*.

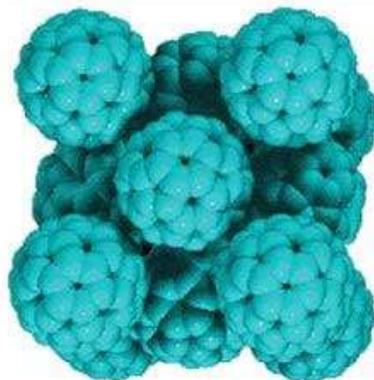


Figure 6.2: A Molecular Crystal. Here, 60 atoms of carbon bind together to form a large molecule known as a buckyball², then the buckyballs can stick together to form a molecular crystal.

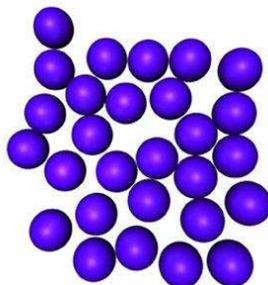


Figure 6.3: Cartoon of a Liquid. In liquids, molecules are not in an ordered configuration and are free to move around (i.e, the liquid can flow). However, the liquid molecules do attract each other and at any moment in time you can typically define neighbors.

Another form of matter is liquid. Here, atoms are attracted to each other, but not so strongly that they form permanent bonds (or the temperature is high enough to make the bonds unstable). Liquids (and gases)³ are disordered configurations of molecules where the molecules are

²The name “buckyball” is an nickname for Buckminsterfullerene, named after Richard Buckminster Fuller, the famed developer of the geodesic dome, which buckyballs are supposed to resemble; although the shape is actually precisely that of a soccer ball. This name is credited to the discoverers of the buckyball, Harold Kroto, James Heath, and Richard Smalley, who were awarded a Nobel prize in chemistry for their discovery despite their choice of nomenclature. (Probably the name “Soccerballene” would have been better).

³As we should have learned in our stat-mech and thermo courses, there is no “fundamental” difference between a liquid and a gas. Generally liquids are high density and not very compressible, whereas gases are low density and very compressible. A single substance (say, water) may have a phase transition between its gas and liquid phase (boiling), but one can also go continuously from the gas to liquid phase without boiling by going to high pressure and going around the critical point (becoming “supercritical”).

free to move around into new configurations.

Somewhere midway between the idea of a crystal and the idea of a liquid is the possibility of amorphous solids and glasses. In this case the atoms are bonded into position in a disordered configuration. Unlike a liquid, the atoms cannot flow freely.

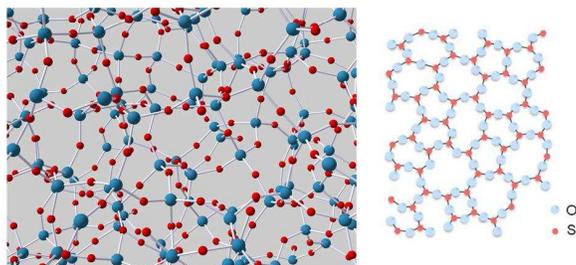


Figure 6.4: Cartoon of Amorphous Solid: Silica (SiO_2) can be an amorphous solid, or a glass (as well as being crystalline quartz). Left is a three dimensional picture, and right is a two dimensional cartoon. Here the atoms are disordered, but are bonded together and cannot flow.

Many more possibilities exist. For example, one may have so-called liquid-crystals, where the system orders in some ways but remains disordered in other ways. For example, in figure 6 the system is crystalline (ordered) in one direction, but remains disordered within each plane. One can also consider cases where the molecules are always oriented the same way but are at completely random positions (known as a “nematic”). There are a huge variety of possible liquid crystal phases of matter. In every case it is the interactions between the molecules (“bonding” of some type, whether it be weak or strong) that dictates the configurations.



Figure 6.5: Cartoon of a Liquid Crystal. Liquid crystals have some of the properties of a solid and some of the properties of a liquid. In this picture of a smectic-C liquid crystal the system is crystalline in the vertical direction (forming discrete layers) but remains liquid (random positions) within each plane. Like a crystal, in this case, the individual molecules all have the same orientation.

One should also be aware of polymers,⁴ which are long chains of atoms (such as DNA).

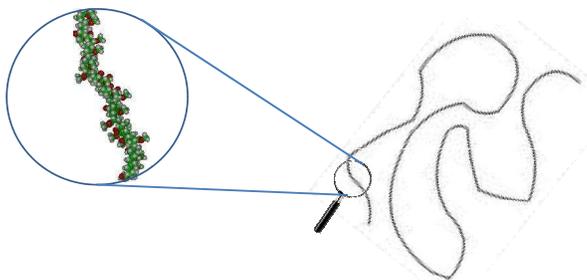


Figure 6.6: Cartoon of a Polymer: A polymer is a long chain of atoms.

And there are many more types of condensed matter systems that we simply do not have time to discuss⁵. One can even engineer artificial types of order which do not occur naturally. Each one of these types of matter has its own interesting properties and if we had more time we would discuss them all in depth! Given that there are so many types of matter, it may seem odd that we are going to spend essentially the entire remainder of our time focused on simple crystalline solids. There are very good reasons for this however. First of all, the study of solids is one of the most successful branches of physics – both in terms of how completely we understand them and also in terms of what we have been able to do practically with this understanding (For example, the entire modern semiconductor industry is a testament to how successful our understanding of solids is). More importantly, however, the physics that we learn by studying solids forms an excellent starting point for trying to understand the many more complex forms of matter that exist.

References

- Dove, chapter 2 gives discussion of many types of matter.

For an even more complete survey of the types of condensed matter see “Principles of Condensed Matter Physics”, by Chaikin and Lubensky (Cambridge).

⁴Here is a really cool experiment to do in your kitchen. Cornstarch is a polymer — a long chain of atoms. Take a box of cornstarch and make a mixture of roughly half cornstarch and half water (you may have to play with the proportions). The concoction should still be able to flow. And if you put your hand into it, it will feel like a liquid and be gooey. But if you take a tub of this and hit it with a hammer very quickly, it will feel as hard as a brick, and it will even crack (then it turns back to goo). In fact, you can make a deep tub of this stuff and although it feels completely like a fluid, you can run across the top of it (If you are too lazy to try doing this try Googling “Ellen cornstarch” to see a YouTube video of the experiment). This mixture is a “non-Newtonian” fluid — its effective viscosity depends on how fast the force is applied to the material. The reason that polymers have this property is that the long polymer strands get tangled with each other. If a force is applied slowly the strands can unentangle and flow past each other. But if the force is applied quickly they cannot unentangle fast enough and the material acts just like a solid.

⁵Particularly interesting are forms such as superfluids, where quantum mechanics dominates the physics. But alas, we must save discussion of this for another course!

Part III

Toy Models of Solids in One Dimension

Chapter 7

One Dimensional Model of Compressibility, Sound, and Thermal Expansion

In the first few chapters we found that our simple models of solids, and electrons in solids, were insufficient in several ways. In order to improve our understanding, we decided that we needed to take the periodic microstructure of crystals more seriously. In this part of the book we finally begin this more careful microscopic consideration. To get a qualitative understanding of the effects of the periodic lattice, it is frequently sufficient to think in terms of simple one dimensional systems. This is our strategy for the next few chapters. Once we have introduced a number of important principles in one dimension, we will address the complications associated with higher dimensionality.

In the last part of the book we discussed bonding between atoms. We found, particularly in the discussion of covalent bonding, that the lowest energy configuration would have the atoms at some optimal distance between (See figure 5.7, for example). Given this shape of the energy as a function of distance between atoms we will be able to come to some interesting conclusions.

For simplicity, let us imagine a one dimensional system of atoms (atoms in a single line). The potential $V(x)$ between two neighboring atoms is drawn in the Figure 7.1.

The classical equilibrium position is the position at the bottom of the well (marked x_{eq} in the figure). The distance between atoms at low temperature should then be x_{eq} . (A good homework assignment is to consider how quantum mechanics can change this value and increase it a little bit!).

Let us now Taylor expand the potential around its minimum.

$$V(x) \approx V(x_{eq}) + \frac{\kappa}{2}(x - x_{eq})^2 + \frac{\kappa_3}{3!}(x - x_{eq})^3 + \dots$$

Note that there is no linear term (if there were a linear term, then the position x_{eq} would not be the minimum). If there are only small deviations from the position x_{eq} the higher terms are much much smaller than the leading quadratic term and we can throw these terms out. This is a rather crucial general principle that *any* potential, close enough to its minimum, is quadratic.

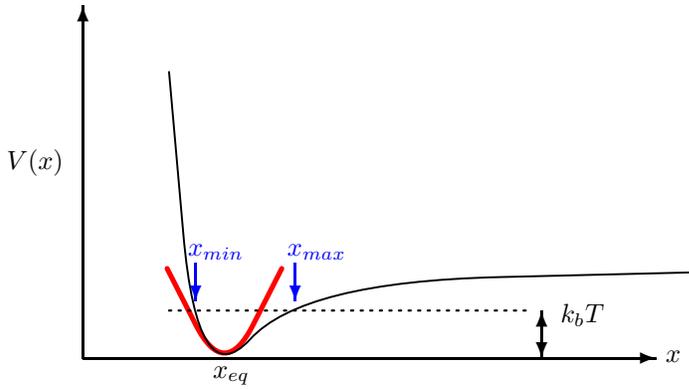


Figure 7.1: Potential Between Neighboring Atoms (black). The thick red curve is a quadratic approximation to the minimum (it may look crooked but in fact the red curve is symmetric and the black curve is asymmetric). The equilibrium position is x_{eq} . At finite temperature T , the system can oscillate between x_{max} and x_{min} which are not symmetric around the minimum. Thus as T increases the *average* position moves out to larger distance and the system expands.

Compressibility (or Elasticity)

We thus have a simple Hooke's law quadratic potential around the minimum. If we apply a force to compress the system (i.e., apply a pressure to our model one dimensional solid) we find

$$-\kappa(\delta x_{eq}) = F$$

where the sign is so that a positive (compressive) pressure reduces the distance between atoms. This is obviously just a description of the compressibility (or elasticity) of a solid. The usual description of compressibility is

$$\beta = -\frac{1}{V} \frac{\partial V}{\partial P}$$

(one should ideally specify if this is measured at fixed T or at fixed S . Here, we are working at $T = S = 0$ for simplicity). In the one dimensional case, we write the compressibility as

$$\beta = -\frac{1}{L} \frac{\partial L}{\partial F} = \frac{1}{\kappa x_{eq}} = \frac{1}{\kappa a} \quad (7.1)$$

with L the length of the system and x_{eq} is the spacing between atoms. Here we make the conventional definition that the equilibrium distance between identical atoms in a system (the so-called *lattice constant*) is written as a .

Sound

You may recall from your fluids course that in an isotropic compressible fluid, one predicts sound waves with velocity

$$v = \sqrt{\frac{B}{\rho}} = \sqrt{\frac{1}{\rho\beta}} \quad (7.2)$$

where ρ is the mass density of the fluid, B is the bulk modulus, which is $B = 1/\beta$ with β the (adiabatic) compressibility.

While in a real solid the compressibility is anisotropic and the speed of sound depends in detail on the direction of propagation, in our model one dimensional solid this is not a problem and we can calculate that the density is m/a with m the mass of each particle and a the equilibrium spacing between particles.

Thus using our result from above, we predict a sound wave with velocity

$$v = \sqrt{\frac{\kappa a^2}{m}} \quad (7.3)$$

Shortly (in section 8.2) we will re-derive this expression from the microscopic equations of motion for the atoms in the one dimensional solid.

Thermal Expansion

So far we have been working at zero temperature, but it is worth thinking at least a little bit about thermal expansion. This will be fleshed out more completely in a homework assignment. (In fact even in the homework assignment the treatment of thermal expansion will be very crude, but that should still be enough to give us the general idea of the phenomenon¹).

Let us consider again figure 7.1 but now at finite temperature. We can imagine the potential as a function of distance between atoms as being like a ball rolling around in a potential. At zero energy, the ball sits at the the minimum of the distribution. But if we give the ball some finite temperature (i.e, some energy) it will oscillate around the minimum. At fixed energy $k_b T$ the ball rolls back and forth between the points x_{min} and x_{max} where $V(x_{min}) = V(x_{max}) = k_b T$. But away from the minimum the potential is asymmetric, so $|x_{max} - x_{eq}| > |x_{min} - x_{eq}|$ so on average the particle has a position $\langle x \rangle > x_{eq}(T = 0)$. This is in essence the reason for thermal expansion! We will obtain positive thermal expansion for any system where $\kappa_3 < 0$ (i.e., at small x the potential is steeper) which almost always is true for real solids.

Summary

- Forces between atoms determine ground state structure.
- These same forces, perturbing around the ground state, determine elasticity, sound velocity, and thermal expansion.
- Thermal expansion comes from the non-quadratic part of the interatomic potential.

Sound and Compressibility:

- Goodstein, section 3.2b
- Ibach and Luth, beginning of section 4.5
- Hook and Hall, section 2.2

Thermal Expansion (Most references go into *way* too much depth on thermal expansion):

- Kittel chapter 5, section on thermal expansion.

¹Although this description is an annoyingly crude discussion of thermal expansion, we are mandated by the IOP to teach something on this subject. Explaining it more correctly is, unfortunately, rather messy!

Chapter 8

Vibrations of a One Dimensional Monatomic Chain

In chapter 2 we considered the Boltzmann, Einstein, and Debye models of vibrations in solids. In this chapter we will consider a detailed model of vibration in a solid, first classically, and then quantum mechanically. We will be able to better understand what these early attempts to understand vibrations achieved and we will be able to better understand their shortcomings.

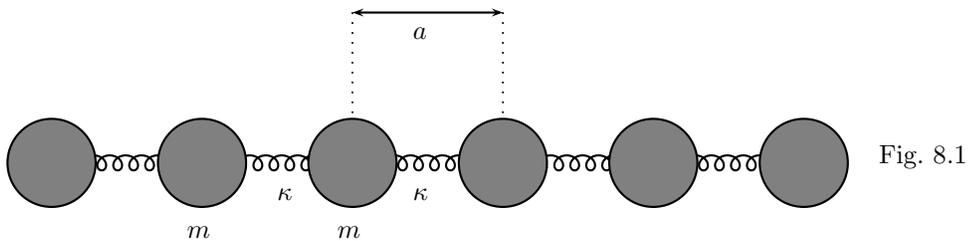
Let us consider a chain of identical atoms of mass m where the equilibrium spacing between atoms is a . Let us define the position of the n^{th} atom to be x_n and the equilibrium position of the n^{th} atom to be $x_n^{\text{eq}} = na$.

Once we allow motion of the atoms, we will have x_n deviating from its equilibrium position, so we define the small variable

$$\delta x_n = x_n - x_n^{\text{eq}}$$

Note that in our simple model we are allowing motion of the masses only in one dimension (i.e., we are allowing longitudinal motion of the chain, not transverse motion).

As discussed in the previous section, if the system is at low enough temperature we can consider the potential holding the atoms together to be quadratic. Thus, our model of a solid is a chain of masses held together with springs as show in this figure



Since the springs are quadratic potentials this model is frequently known as a *harmonic* chain.

With this quadratic interatomic potential, we can write the total potential energy of the chain to be

$$\begin{aligned} V_{tot} &= \sum_i V(x_i - x_{i+1}) \\ &= V_{eq} + \sum_i \frac{\kappa}{2} (\delta x_i - \delta x_{i+1})^2 \end{aligned}$$

The force on the n^{th} mass on the chain is then given by

$$F_n = -\frac{\partial V_{tot}}{\partial x_n} = \kappa(\delta x_{n+1} - \delta x_n) + \kappa(\delta x_{n-1} - \delta x_n)$$

Thus we have Newton's equation of motion

$$m(\delta \ddot{x}_n) = F_n = \kappa(\delta x_{n+1} + \delta x_{n-1} - 2\delta x_n) \quad (8.1)$$

To remind the reader, for any coupled system, a *normal mode* is defined to be a collective oscillation where all particles move at the same frequency. We now attempt a solution to Newton's equations by using an ansatz that describes the normal modes as waves

$$\delta x_n = Ae^{i\omega t - ikx_n^{eq}} = Ae^{i\omega t - ikna}$$

where A is an amplitude of oscillation.

Now the reader might be confused about how it is that we are considering complex values of δx_n . Here we are using complex numbers for convenience but actually we implicitly mean to take the real part. (This is analogous to what one does in circuit theory with oscillating currents!). Since we are taking the real part, it is sufficient to consider only $\omega \geq 0$, however, we must be careful that k can then have either sign, and these are inequivalent once we have specified that ω is positive.

Plugging our ansatz into Eq. 8.1 we obtain

$$-m\omega^2 Ae^{i\omega t - ikna} = \kappa Ae^{i\omega t} \left[e^{-ika(n+1)} + e^{-ika(n-1)} - 2e^{-ikan} \right]$$

or

$$m\omega^2 = 2\kappa[1 - \cos(ka)] = 4\kappa \sin^2(ka/2) \quad (8.2)$$

We thus obtain the result

$$\omega = 2\sqrt{\frac{\kappa}{m}} \left| \sin\left(\frac{ka}{2}\right) \right| \quad (8.3)$$

In general a relationship between a frequency (or energy) and a wavevector (or momentum) is known as a *dispersion relation*. This particular dispersion relation is shown in Fig. 8.1

8.1 First Exposure to the Reciprocal Lattice

Note that in Fig. 8.1 we have only plotted the dispersion for $-\pi/a \leq k \leq \pi/a$. The reason for this is obvious from Eq. 8.3 — the dispersion relation is actually periodic in $k \rightarrow k + 2\pi/a$. In fact this is a very important general principle:

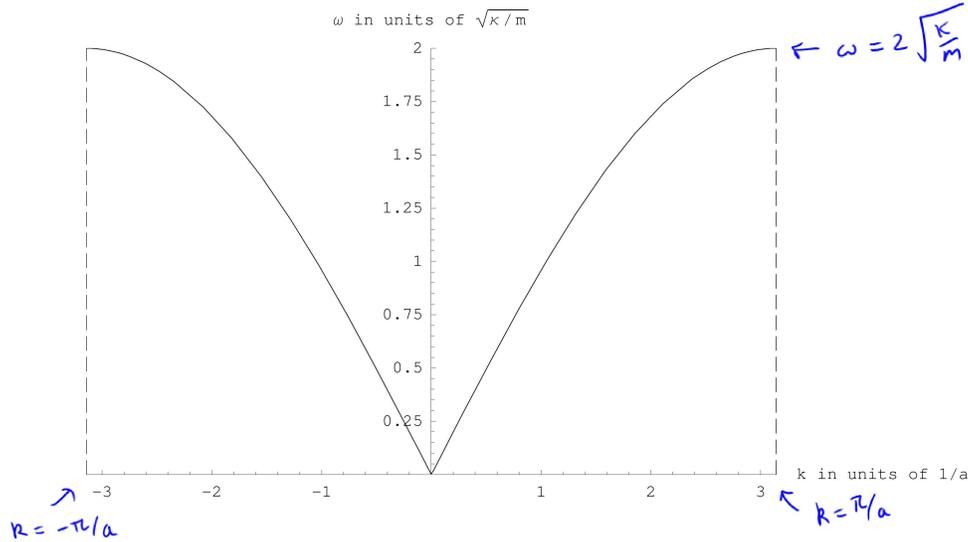


Figure 8.1: Dispersion Relation for Vibrations of the One Dimensional Monatomic Harmonic Chain. The dispersion is periodic in $k \rightarrow k + 2\pi/a$

Principle 8.1: A system which is periodic in real space with a periodicity a will be periodic in reciprocal space with periodicity $2\pi/a$.

In this principle we have used the word *reciprocal* space which means k -space. In other words this principle tells us that if a system looks the same when $x \rightarrow x + a$ then in k -space the dispersion will look the same when $k \rightarrow k + 2\pi/a$. We will return to this principle many times in later chapters.

The periodic unit (the “unit cell”) in k -space is conventionally known as the *Brillouin Zone*^{1,2}. This is your first exposure to the concept of a Brillouin zone, but it will play a very central role in later chapters. The “First Brillouin Zone” is a unit cell in k -space centered around the point $k = 0$. Thus in Fig. 8.1 we have shown only the first Brillouin zone, with the understanding that the dispersion is periodic for higher k . The points $k = \pm\pi/a$ are known as the *Brillouin-Zone boundary* and are defined in this case as being points which are symmetric around $k = 0$ and are separated by $2\pi/a$.

It is worth pausing for a second and asking why we expect that the dispersion curve should

¹Leon Brillouin was one of Sommerfeld’s students. He is famous for many things including for being the “B” in the “WKB” approximation. I’m not sure if WKB is on your syllabus, but it really should be if it is not already!

²The pronunciation of “Brillouin” is something that gives English speakers a great deal of difficulty. If you speak French you will probably cringe at the way this name is butchered. (I did badly in French in school, so I’m probably one of the worst offenders.) According to online dictionaries it is properly pronounced somewhere between the following words: brēwan, breel-wahn, bree(y)lwa(n), and bree-l(uh)-wahn. At any rate, the “l” and the “n” should both be very weak.

be periodic in $k \rightarrow k + 2\pi/a$. Recall that we defined our vibration mode to be of the form

$$\delta x_n = Ae^{i\omega t - ikna} \quad (8.4)$$

If we take $k \rightarrow k + 2\pi/a$ we obtain

$$\delta x_n = Ae^{i\omega t - i(k+2\pi/a)na} = Ae^{i\omega t - ikna} e^{-i2\pi n} = Ae^{i\omega t - ikna}$$

where here we have used

$$e^{-i2\pi n} = 1$$

for any integer n . What we have found here is that shifting $k \rightarrow k + 2\pi/a$ gives us back exactly the same oscillation mode the we had before we shifted k . The two are physically exactly equivalent!

In fact, it is similarly clear that shifting k by any $k + 2\pi p/a$ with p an integer will give us back exactly the same wave also since

$$e^{-i2\pi np} = 1$$

as well. We can thus define a set of points in k -space (reciprocal space) which are all physically equivalent to the point $k = 0$. This set of points is known as the *reciprocal lattice*. The original periodic set of points $x_n = na$ is known as the *direct lattice* or *real-space lattice* to distinguish it from the reciprocal lattice, when necessary.

The concept of the reciprocal lattice will be extremely important later on. We can see the analogy between the direct lattice and the reciprocal lattice as follows:

$$\begin{array}{cccccccc} x_n & = & \dots & -2a, & -a, & 0, & a, & 2a, & \dots \\ G_n & = & \dots & -2\left(\frac{2\pi}{a}\right), & -\frac{2\pi}{a}, & 0, & \frac{2\pi}{a}, & 2\left(\frac{2\pi}{a}\right), & \dots \end{array}$$

Note that the defining property of the reciprocal lattice in terms of the points in the real lattice can be given as

$$e^{iG_n x_n} = 1 \quad (8.5)$$

A point G_m is a member of the reciprocal lattice if and only if Eq. 8.5 is true for all x_n in the real lattice.

8.2 Properties of the Dispersion of the One Dimensional Chain

We now return to more carefully examine the properties of the dispersion we calculated (Eq. 8.3).

Sound Waves:

Recall that sound wave³ is a vibration that has a long wavelength (compared to the inter-atomic spacing). In this long wavelength regime, we find the dispersion we just calculated to be linear with wavevector $\omega = v_{sound}k$ as expected for sound with

$$v_{sound} = a\sqrt{\frac{\kappa}{m}}.$$

³For reference it is good to remember that humans can hear sound wavelengths roughly between 1cm and 10m. Both of these are very long wavelength compared to interatomic spacings.

(To see this, just expand the sin in Eq. 8.3). Note that this sound velocity matches the velocity predicted from Eq. 7.3!

However, we note that at larger k , the dispersion is no longer linear. This is in disagreement with what Debye assumed in his calculation in section 2.2. So clearly this is a shortcoming of the Debye theory. In reality the dispersion of normal modes of vibration is linear only at long wavelength.

At shorter wavelength (larger k) one typically defines two different velocities: The *group velocity*, the speed at which a wavepacket moves, is given by

$$v_{group} = d\omega/dk.$$

And the *phase velocity*, the speed at which the individual maxima and minima move, is given by

$$v_{phase} = \omega/k.$$

These two match in the case of a linear dispersion, but otherwise are different. Note that the group velocity becomes zero at the Brillouin zone boundaries $k = \pm\pi/a$ (i.e., the dispersion is flat). As we will see many times later on, this is a general principle!

Counting Normal Modes:

Let us now ask how many normal modes there are in our system. Naively it would appear that we can put any k such that $-\pi/a \leq k < \pi/a$ into Eq. 8.3 and obtain a new normal mode with wavevector k and frequency $\omega(k)$. However this is not precisely correct.

Let us assume our system has exactly N masses in a row, and for simplicity let us assume that our system has periodic boundary conditions i.e., particle x_0 has particle x_1 to its right and particle x_{N-1} to its left. Another way to say this is to let, $x_{n+N} = x_n$, i.e., this one dimensional system forms a big circle. In this case we must be careful that the wave ansatz Eq. 8.4 makes sense as we go all the way around the circle. We must therefore have

$$e^{i\omega t - ikna} = e^{i\omega t - ik(N+n)a}$$

Or equivalently we must have

$$e^{ikNa} = 1$$

This requirement restricts the possible values of k to be of the form

$$k = \frac{2\pi p}{Na} = \frac{2\pi p}{L}$$

where p is an integer and L is the total length of the system. Thus k becomes quantized rather than a continuous variable. This means that the k -axis in Figure 8.1 is actually a discrete set of many many individual points; the spacing between two of these consecutive points being $2\pi/(Na) = 2\pi/L$.

Let us now count how many normal modes we have. As mentioned above in our discussion of the Brillouin zone, adding $2\pi/a$ to k brings one back to exactly the same physical wave. Thus we only ever need consider k values within the first Brillouin zone (i.e., $-\pi/a \leq k < \pi/a$, and since π/a is the same as $-\pi/a$ we choose to count one but not the other). Thus the total number of normal modes is

$$\text{Total Number of Modes} = \frac{\text{Range of } k}{\text{Spacing between neighboring } k} = \frac{2\pi/a}{2\pi/(Na)} = N. \quad (8.6)$$

There is precisely one normal mode per mass in the system — that is, one normal mode per degree of freedom in the whole system. This is what Debye insightfully predicted in order to cut off his divergent integrals in section 2.2.3 above!

8.3 Quantum Modes: Phonons

We now make a rather important leap from classical to quantum physics.

Quantum Correspondence: If a classical harmonic system (i.e., any quadratic Hamiltonian) has a normal oscillation mode at frequency ω the corresponding quantum system will have eigenstates with energy

$$E_n = \hbar\omega\left(n + \frac{1}{2}\right) \quad (8.7)$$

Presumably you know this well in the case of a single harmonic oscillator. The only thing different here is that our harmonic oscillator can be a collective normal mode not just motion of a single particle. This quantum correspondence principle will be the subject of a homework assignment.

Thus at a given wavevector k , there are many possible eigenstates, the ground state being the $n = 0$ eigenstate which has only the zero-point energy $\hbar\omega(k)/2$. The lowest energy excitation is of energy $\hbar\omega(k)$ greater than the ground state corresponding to the excited $n = 1$ eigenstate. Generally all excitations at this wavevector occur in energy units of $\hbar\omega(k)$, and the higher values of energy correspond classically to oscillations of increasing amplitude.

Each excitation of this “normal mode” by a step up the harmonic oscillator excitation ladder (increasing the quantum number n) is known as a “phonon”.

Definition 8.3.1. A *phonon* is a discrete quantum of vibration⁴

This is entirely analogous to defining a single quanta of light as a photon. As is the case with the photon, we may think of the phonon as actually being a particle, or we can think of the phonon as being a quantized wave.

If we think about the phonon as being a particle (as with the photon) then we see that we can put many phonons in the same state (ie., the quantum number n in Eq. 8.7 can be increased to any value), thus we conclude that phonons, like photons, are bosons. As with photons, at finite temperature there will be a nonzero number of phonons (i.e., n will be on average nonzero) as given by the Bose occupation factor.

$$n_B(\beta\hbar\omega) = \frac{1}{e^{\beta\hbar\omega} - 1}$$

with $\beta = 1/(k_bT)$ and ω the oscillation frequency.

Thus, the energy expectation of the phonons at wavevector k is given by

$$E_k = \hbar\omega(k) \left(n_B(\beta\hbar\omega(k)) + \frac{1}{2} \right)$$

⁴I do not like the definition of a phonon as “a quantum of vibrational energy” which many books use. The vibration does carry indeed energy, but it carries other quantum numbers (such as crystal momentum) as well, so why specify energy only?

We can use this type of expression to calculate the heat capacity of our one dimensional model⁵

$$U_{total} = \sum_k \hbar\omega(k) \left(n_B(\beta\hbar\omega(k)) + \frac{1}{2} \right)$$

where the sum over k here is over all possible normal modes, i.e, $k = 2\pi p/(Na)$ such that $-\pi/a \leq k < \pi/a$. Thus we really mean

$$\sum_k \rightarrow \sum_{\substack{p=-N/2 \\ k=(2\pi p)/(Na)}}^{p=N/2-1}$$

Since for a large system, the k points are very close together, we can convert the discrete sum into an integral (something we should be very familiar with by now) to obtain

$$\sum_k \rightarrow \frac{Na}{2\pi} \int_{-\pi/a}^{\pi/a} dk$$

Note that we can use this continuum integral to count the total number of modes in the system

$$\frac{Na}{2\pi} \int_{-\pi/a}^{\pi/a} dk = N$$

as predicted by Debye.

Using this integral form of the sum, we have the total energy given by

$$U_{total} = \frac{N}{2\pi} \int_{-\pi/a}^{\pi/a} dk \hbar\omega(k) \left(n_B(\beta\hbar\omega(k)) + \frac{1}{2} \right)$$

from this we could calculate specific heat as dU/dT .

These two previous expressions look exactly like what Debye would have obtained from his calculation (for a one dimensional version of his model)! The only difference lies in our expression for $\omega(k)$. Debye only knew about sound where $\omega = vk$, is linear in the wavevector. We, on the other hand, have just calculated that for our microscopic ball and spring model ω is not linear in k (See Eq. 8.3). Other than this change in the dispersion relation, our calculation of heat capacity (exact for this model!) is identical to the approach of Debye. In fact, Einstein's calculation of specific heat can also be phrased in exactly the same language. Only for Einstein's model the frequency ω is constant for all k (it is fixed at the Einstein frequency). We thus see Einstein's model, Debye's model, and our microscopic harmonic model in a very unified light. The only difference between the three is what we use for a dispersion relation.

One final comment is that it is frequently useful to further replace integrals over k with integrals over frequency (we did this when we studied the Debye model above). We obtain generally

$$\frac{Na}{2\pi} \int_{-\pi/a}^{\pi/a} dk = \int d\omega g(\omega)$$

where⁶

⁵The observant reader will note that we are calculating $C_V = dU/dT$ the heat capacity at constant volume. Why constant volume? As we saw above when we studied thermal expansion, the crystal does not expand unless we include third(or higher) order terms in the interatomic potential, which are not in this model!

⁶The factor of 2 out front comes from the fact that each ω occurs for the two possible values of $\pm k$.

$$g(\omega) = 2 \frac{Na}{2\pi} |dk/d\omega|$$

Recall again that the definition of density of states is that the number of modes with frequency between ω and $\omega + d\omega$ is given by $g(\omega)d\omega$.

Note that in the (one dimensional) Debye model this density of states is constant from $\omega = 0$ to $\omega = \omega_{Debye} = v\pi/a$. In our model, as we have calculated above, the density of states is not a constant, but becomes zero at frequency above the maximum frequency $2\sqrt{\kappa/m}$. (In a homework problem we calculate this density of states explicitly). Finally in the Einstein model, this density of states is a delta-function at the Einstein frequency.

8.4 Crystal Momentum

As mentioned above, the wavevector of a phonon is defined only modulo⁷ the reciprocal lattice. In other words, k is the same as $k + G_m$ where $G_m = 2\pi m/a$ is a point in the reciprocal lattice. Now we are supposed to think of these phonons as particles — and we like to think of our particles as having energy $\hbar\omega$ and a momentum $\hbar k$. But we cannot define a phonon's momentum this way because physically it is the same phonon whether we describe it as $\hbar k$ or $\hbar(k + G_m)$. We thus instead define a concept known as the *crystal momentum* which is the momentum modulo the reciprocal lattice — or equivalently we agree that we must always describe k within the first Brillouin zone.

In fact, this idea of crystal momentum is extremely powerful. Since we are thinking about phonons as being particles, it is actually possible for two (or more) phonons to bump into each other and scatter from each other — the same way particles do⁸. In such a collision, energy is conserved and *crystal momentum is conserved!* For example three phonons each with crystal momentum $\hbar(2/3)\pi/a$ can scatter off of each other to produce three phonons each with crystal momentum $-\hbar(2/3)\pi/a$. This is allowed since the initial and final states have the same energy and

$$3 \times (2/3)\pi/a = 3 \times (-2/3)\pi/a \pmod{2\pi/a}$$

During these collisions although momentum $\hbar k$ is not conserved, crystal momentum is⁹. In fact, the situation is similar when, for example, phonons scatter from electrons in a periodic lattice — crystal momentum becomes the conserved quantity rather than momentum. This is an extremely important principle which we will encounter again and again. In fact, it is a main cornerstone of solid-state physics.

Aside: There is a very fundamental reason for the conservation of crystal momentum. Conserved

⁷The word “modulo” or “mod” means to divide and only keep the remainder. For example, 15 modulo 7 = 1 since when you divide 15 by 7, you have a remainder of 1.

⁸In the harmonic model we have considered phonons do not scatter from each other. We know this because the phonons are eigenstates of the system, so their occupation does not change with time. However, if we add anharmonic (cubic and higher) terms to the inter-atomic potential, this corresponds to perturbing the phonon Hamiltonian and can be interpreted as allowing phonons to scatter from each other.

⁹This thing we have defined, $\hbar k$, has dimensions of momentum, but is not conserved. However, as we will discuss below in chapter 13, if a particle, like a photon, enters a crystal with a given momentum and undergoes a process that conserves crystal momentum but not momentum, when the photon exits the crystal we will find that total momentum of the system is indeed conserved, with the momentum of the entire crystal accounting for any momentum that is missing from the photon. See footnote 6 in section 13.1.1

quantities are results of symmetries (this is a deep and general statement known as Noether's theorem¹⁰). For example, conservation of momentum is a result of the translational invariance of space. If space is not the same from point to point, for example if there is a potential $V(x)$ which is different at different places, then momentum is not conserved. The conservation of crystal momentum correspondingly results from space being invariant under translations of a , giving us momentum that is conserved modulo $2\pi/a$.

8.5 Summary of Vibrations of the One Dimensional Monatomic Chain

A number of very crucial new ideas have been introduced in this section. Many of these will return again and again in later chapters.

- Normal modes are collective oscillations where all particles move at the same frequency.
- If a system is periodic in space with periodicity $\Delta x = a$, then in reciprocal space (k -space) the system is periodic with periodicity $\Delta k = 2\pi/a$.
- Values of k which differ by multiples of $2\pi/a$ are physically equivalent. The set of points in k -space which are equivalent to $k = 0$ are known as the reciprocal lattice.
- Any value of k is equivalent to some k in the first Brillouin-zone, $-\pi/a \leq k < \pi/a$ (in 1d).
- The sound velocity is the slope of the dispersion in the small k limit (group = phase velocity in this limit).
- A classical normal mode of frequency ω gets translated into quantum mechanical eigenstates $E_n = \hbar\omega(n + \frac{1}{2})$. If the system is in the n^{th} eigenstate, we say that it is occupied by n phonons.
- Phonons can be thought of as particles, like photons, that obey Bose statistics.

References

Normal Modes of Monatomic Chain and Introduction to Phonons:

- Kittel, beginning of chapter 4
- Goodstein, beginning of section 3.3
- Hook and Hall, section 2.3.1
- Burns, section 12.1–12.2
- Ashcroft and Mermin, beginning of chapter 22.

¹⁰Emmy Noether has been described by Einstein, among others, as the most important woman in the history of mathematics.

Chapter 9

Vibrations of a One Dimensional Diatomic Chain

In the previous chapter we studied in detail a one dimensional model of a solid where every atom is identical to every other atom. However, in real materials not every atom is the same (for example, in sodium chloride, NaCl, we have two types of atoms!). We thus intend to generalize our previous discussion of the one dimension solid to a one dimensional solid with two types of atoms. Much of this will follow the outline set in the previous chapter, but we will see that several fundamentally new features will now emerge.

9.1 Diatomic Crystal Structure: Some useful definitions

Consider the following model system

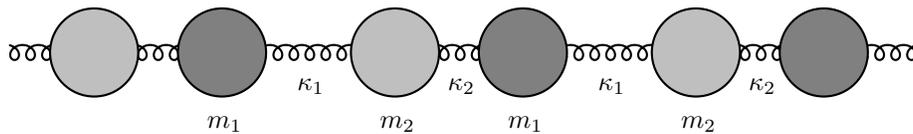


Fig. 9.1.1

which represents a periodic arrangement of two different types of atoms. Here we have given them two masses m_1 and m_2 which alternate along the one dimensional chain. The springs connecting the atoms have spring constants κ_1 and κ_2 and also alternate.

In this circumstance with more than one type of atom, we first would like to identify the so-called *unit cell* which is the repeated motif in the arrangement of atoms. In this picture, we have put a box around the unit cell. The length of the unit cell in one dimension is known as the

lattice constant and it is labeled a .

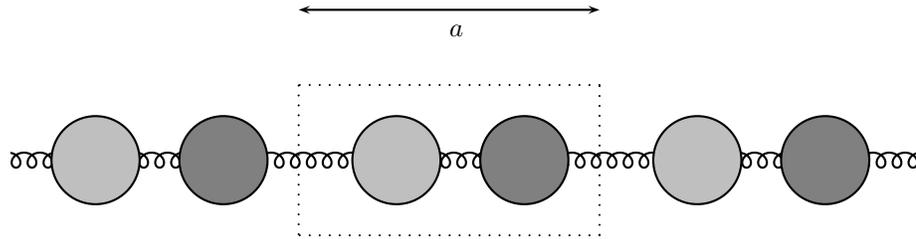


Fig. 9.1.2

Note however, that the definition of the unit cell is extremely non-unique. We could just as well have chosen (for example) the unit cell to be as follows.

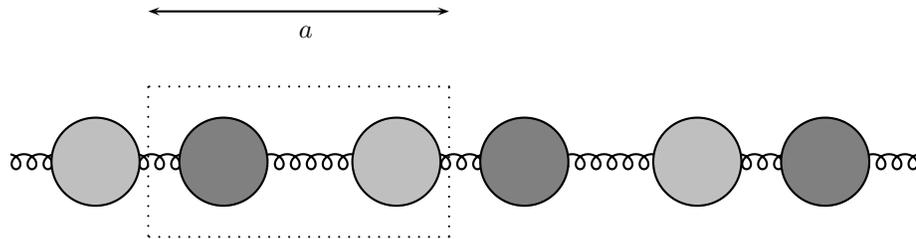


Fig. 9.1.3

The important thing in defining a periodic system is to choose *some* unit cell and then construct the full system by reproducing the same unit cell over and over. (In other words, make a definition of the unit cell and stick with that definition!).

It is sometimes useful to pick some reference point inside each unit cell. This set of reference points makes a simple *lattice* (we will define the term “lattice” more closely in later chapters, but for now the point is that a lattice has only one type of point in it – not two different types of points). So in this figure, we have marked our reference point in each unit cell with an X (again, the choice of this reference point is arbitrary).

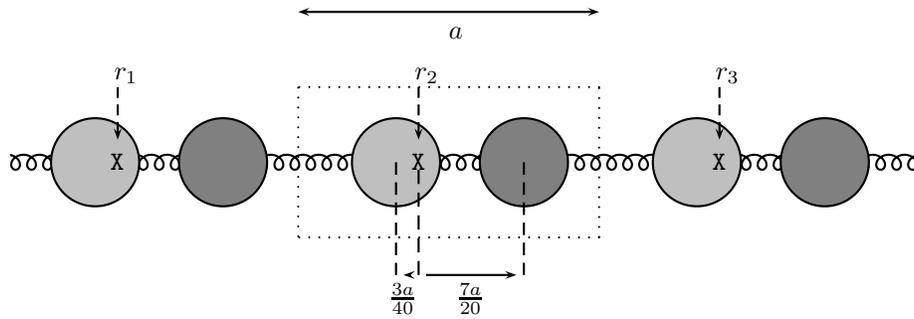


Fig. 9.1.4

Given the reference lattice point in the unit cell, the description of all of the atoms in the unit cell with respect to this reference point is known as a *basis*. In this case we might describe our basis as

- light gray atom centered at position $3a/40$ to left of reference lattice point
- dark gray atom centered at position $7a/20$ to right of reference lattice point

Thus if the reference lattice point in unit cell n is called r_n (and the spacing between the lattice points is a) we can set

$$r_n = an$$

with a the size of the unit cell. Then the (equilibrium) position of the light gray atom in the n^{th} unit cell is

$$x_n^{eq} = an - 3a/40$$

whereas the (equilibrium) position of the dark gray atom in the n^{th} unit cell is

$$y_n^{eq} = an + 7a/20$$

9.2 Normal Modes of the Diatomic Solid

For simplicity, let us focus on the case where all of the masses along our chain are the same $m_1 = m_2 = m$ but the two spring constants κ_1 and κ_2 are different. (For homework we will consider the case where the masses are different, but the spring constants are the same!).

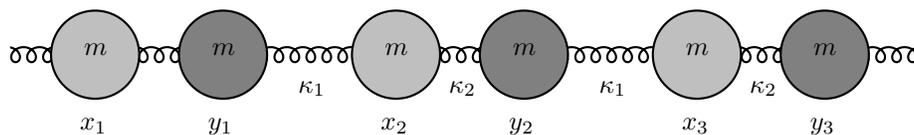


Fig. 9.2.1

Given the spring constants in the picture, we can write down Newton's equations of motion for the deviations of the positions of the masses from their equilibrium positions. We obtain

$$m \ddot{\delta x}_n = \kappa_2(\delta y_n - \delta x_n) + \kappa_1(\delta y_{n-1} - \delta x_n) \quad (9.1)$$

$$m \ddot{\delta y}_n = \kappa_1(\delta x_{n+1} - \delta y_n) + \kappa_2(\delta x_n - \delta y_n) \quad (9.2)$$

Analogous to the one dimensional case we propose ansätze¹ for these quantities that have the form of a wave

$$\delta x_n = A_x e^{i\omega t - ikna} \quad (9.3)$$

$$\delta y_n = A_y e^{i\omega t - ikna} \quad (9.4)$$

where, as in the previous chapter, we implicitly mean to take the real part of the complex number. As such, we can always choose to take $\omega > 0$ as long as we consider k to be either positive and negative.

As we saw in the previous chapter, values of k that differ by $2\pi/a$ are physically equivalent. We can thus focus our attention to the first Brillouin zone $-\pi/a \leq k < \pi/a$. Note that the important length here is the unit cell length or lattice constant a . Any k outside the first Brillouin zone is redundant with some other k inside the zone.

As we found in the previous chapter, if our system has N unit cells (hence $L = Na$) then (putting periodic boundary conditions on the system) k will be quantized in units of $2\pi/(Na) = 2\pi/L$. Note that here the important quantity is N , the number of unit cells, not the number of atoms ($2N$).

Dividing the range of k in the first Brillouin zone by the spacing between neighboring k 's, we obtain exactly N different possible values of k exactly as we did in Eq. 8.6. In other words, we have exactly one value of k per unit cell.

We might recall at this point the intuition that Debye used — that there should be exactly one possible excitation mode per degree of freedom of the system. Here we obviously have two degrees of freedom per unit cell, but we obtain only one possible value of k per unit cell. The resolution, as we will see in a moment, is that there will be two possible oscillation modes for each wavevector k .

We now proceed by plugging in our ansätze (Eq. 9.3 and 9.4) into our equations of motion (Eq. 9.1 and 9.2). We obtain

$$\begin{aligned} -\omega^2 m A_x e^{i\omega t - ikna} &= \kappa_2 A_y e^{i\omega t - ikna} + \kappa_1 A_y e^{i\omega t - ik(n-1)a} - (\kappa_1 + \kappa_2) A_x e^{i\omega t - ikna} \\ -\omega^2 m A_y e^{i\omega t - ikna} &= \kappa_1 A_x e^{i\omega t - ik(n+1)a} + \kappa_2 A_x e^{i\omega t - ikna} - (\kappa_1 + \kappa_2) A_y e^{i\omega t - ikna} \end{aligned}$$

which simplifies to

$$\begin{aligned} -\omega^2 m A_x &= \kappa_2 A_y + \kappa_1 A_y e^{ika} - (\kappa_1 + \kappa_2) A_x \\ -\omega^2 m A_y &= \kappa_1 A_x e^{-ika} + \kappa_2 A_x - (\kappa_1 + \kappa_2) A_y \end{aligned}$$

This can be rewritten conveniently as an eigenvalue equation

$$m\omega^2 \begin{pmatrix} A_x \\ A_y \end{pmatrix} = \begin{pmatrix} (\kappa_1 + \kappa_2) & -\kappa_2 - \kappa_1 e^{ika} \\ -\kappa_2 - \kappa_1 e^{-ika} & (\kappa_1 + \kappa_2) \end{pmatrix} \begin{pmatrix} A_x \\ A_y \end{pmatrix} \quad (9.5)$$

¹I believe this is the proper pluralization of ansatz.

The solutions of this are obtained by finding the zeros of the secular determinant

$$0 = \begin{vmatrix} (\kappa_1 + \kappa_2) - m\omega^2 & -\kappa_2 - \kappa_1 e^{ika} \\ -\kappa_2 - \kappa_1 e^{-ika} & (\kappa_1 + \kappa_2) - m\omega^2 \end{vmatrix} = |(\kappa_1 + \kappa_2) - m\omega^2|^2 - |\kappa_2 + \kappa_1 e^{ika}|^2$$

The roots of which are clearly given by

$$m\omega^2 = (\kappa_1 + \kappa_2) \pm |\kappa_1 + \kappa_2 e^{ika}|$$

The second term needs to be simplified

$$|\kappa_1 + \kappa_2 e^{ika}| = \sqrt{(\kappa_1 + \kappa_2 e^{ika})(\kappa_1 + \kappa_2 e^{-ika})} = \sqrt{\kappa_1^2 + \kappa_2^2 + 2\kappa_1\kappa_2 \cos(ka)}$$

So we finally obtain

$$\omega_{\pm} = \sqrt{\frac{\kappa_1 + \kappa_2}{m} \pm \frac{1}{m} \sqrt{\kappa_1^2 + \kappa_2^2 + 2\kappa_1\kappa_2 \cos(ka)}} \quad (9.6)$$

Note in particular that for each k we find two normal modes — usually referred to as the two *branches* of the dispersion. Thus since there are N different k values, we obtain $2N$ modes total (if there are N unit cells in the entire system). This is in agreement with our above discussion that we should have exactly one normal mode per degree of freedom in our system.

The dispersion of these two modes is shown in Figure 9.1.

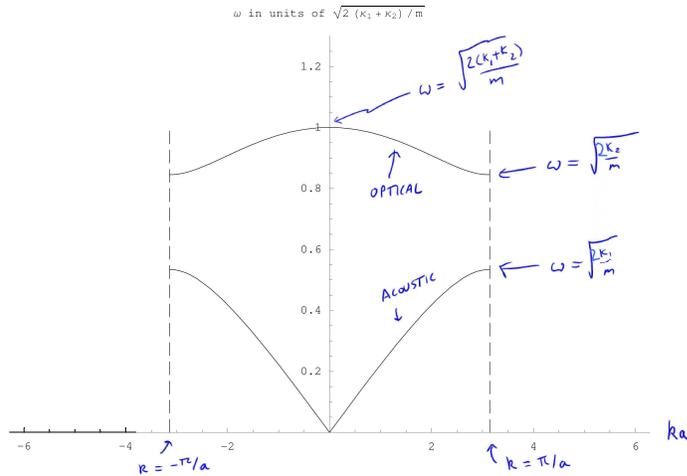


Figure 9.1: Dispersion Relation for Vibrations of the One Dimensional Diatomic Chain. The dispersion is periodic in $k \rightarrow k + 2\pi/a$. Here the dispersion is shown for the case of $\kappa_2 = 1.4\kappa_1$. This scheme of plotting dispersions, putting all normal modes within the first Brillouin zone, is the *reduced zone scheme*. Compare this to Fig. 9.2 below.

A few things to note about this dispersion. First of all we note that there is a long wavelength low energy branch of excitations with linear dispersion (corresponding to ω_- in Eq. 9.6). This is the sound wave, or acoustic mode. Generally the definition of an *acoustic mode* is any mode that has linear dispersion as $k \rightarrow 0$.

By expanding Eq. 9.6 for small k it is easy to check that the sound velocity is

$$v_{\text{sound}} = \frac{d\omega_-}{dk} = \sqrt{\frac{a^2\kappa_1\kappa_2}{2m(\kappa_1 + \kappa_2)}} \quad (9.7)$$

In fact, we could have calculated this sound velocity on general principles analogous to what we did in Eq. 7.2 and Eq. 7.3. The density of the chain is $2m/a$. The effective spring constant of two springs κ_1 and κ_2 in series is $\tilde{\kappa} = (\kappa_1\kappa_2)/(\kappa_1 + \kappa_2)$ so the compressibility of the chain is $\beta = 1/(\tilde{\kappa}a)$ (See Eq. 7.1). Then plugging into Eq. 7.2 gives exactly the same sound velocity as we calculate here in Eq. 9.7.

The higher energy branch of excitations is known as the *optical* mode. It is easy to check that in this case the optical mode goes to frequency $\sqrt{2(\kappa_1 + \kappa_2)}/m$ at $k = 0$, and also has zero group velocity at $k = 0$. The reason for the nomenclature “optical” will become clearer later in the course when we study scattering of light from solids. For now we give a very simplified description of why it is named this way: Consider a solid being exposed to light. It is possible for the light to be absorbed by the solid, but energy and momentum must both be conserved. However, light travels at a very high velocity c , so $\omega = ck$ is a very large number. Since phonons have a maximum frequency, this means that photons can only be absorbed for very small k . However, for small k , acoustic phonons have energy $vk \ll ck$ so that energy and momentum cannot be conserved. On the other hand, optical phonons have energy ω_{optical} which is finite for small k so that at some value of small k , we have $\omega_{\text{optical}} = ck$ and one can match the energy and momentum of the photon to that of the phonon.² Thus, whenever phonons interact with light, it is inevitably the optical phonons that are involved.

Let us examine a bit more closely the acoustic and the optical mode as $k \rightarrow 0$. Examining our eigenvalue problem Eq. 9.5, we see that in this limit the matrix to be diagonalized takes the simple form

$$\omega^2 \begin{pmatrix} A_x \\ A_y \end{pmatrix} = \frac{\kappa_1 + \kappa_2}{m} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} A_x \\ A_y \end{pmatrix} \quad (9.8)$$

The acoustic mode (which has frequency 0) is solved by the eigenvector

$$\begin{pmatrix} A_x \\ A_y \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

This tells us that the two masses in the unit cell (at positions x and y) move together for the case of the acoustic mode in the long wavelength limit. This is not surprising considering our understanding of sound waves as being very long wavelength compressions and rarefactions. This is depicted in Figure 9.2.2. Note in the figure that the amplitude of the compression is slowly modulated, but always the two atoms in the unit cell move almost exactly the same way.

²From this naive argument, one might think that the process where one photon with frequency ω_{optical} is absorbed while emitting a phonon is an allowed process. This is not true since the photons carry spin and spin must also be conserved. Much more typically the interaction between photons and phonons is one where a photon is absorbed and then re-emitted at a different frequency while emitting a phonon. I.e., the photon is inelastically scattered. We will discuss this later on.

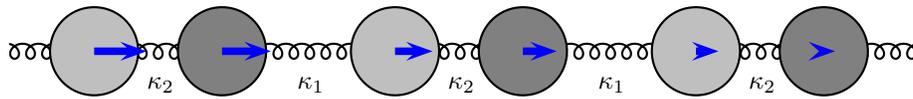


Fig. 9.2.2

A long wavelength acoustic mode

On the other hand, the optical mode at $k = 0$, having frequency $\omega^2 = \frac{2(\kappa_1 + \kappa_2)}{m}$, has the eigenvector

$$\begin{pmatrix} A_x \\ A_y \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

which describes the two masses in the unit cell moving in *opposite* directions, for the optical mode. This is depicted in Figure 9.2.3. Note in the figure that the amplitude of the compression is slowly modulated, but always the two atoms in the unit cell move almost exactly the opposite way.

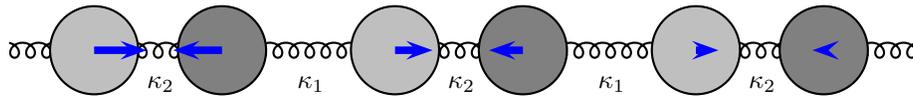


Fig. 9.2.3

A long wavelength optical mode

In order to get a better idea of how motion occurs for both the optical and acoustic modes, it is useful to see animations, which you can find on the web. Another good resource is to download the program “ChainPlot” from Professor Mike Glazer’s web site (<http://www.amg122.com/programs>)³

In this example we had two atoms per unit cell and we obtained two modes per distinct value of k . One of these modes is acoustic and one is optical. More generally if there are M atoms per unit cell (in one dimension) we will have M modes per distinct value of k (i.e., M branches of the dispersion) of which one mode will be acoustic (goes to zero energy at $k = 0$) and all of the remaining modes are optical (do not go to zero energy at $k = 0$).

Caution: We have been careful to discuss a true one dimensional system, where the atoms are allowed to move only along the one dimensional line. Thus each atom has only one degree of freedom. However, if we allow atoms to move in other directions (transverse to the 1d line) we will have more degrees of freedom per atom. When we get to the 3d solid we should expect 3 degrees of freedom per atom. And there should be 3 different acoustic modes at each k at long wavelength. (In 3d, if there are n atoms per unit cell, there will be $3(n - 1)$ optical modes but always 3 acoustic modes totalling $3n$ degrees of freedom per unit cell.

³Note in particular the comment on this website about most books getting the form of the acoustic mode incorrect!

One thing that we should study closely is the behavior at the Brillouin zone boundary. It is also easy to check that the frequencies ω_{\pm} at the zone boundary ($k = \pm\pi/a$) are $\sqrt{2\kappa_1/m}$ and $\sqrt{2\kappa_2/m}$, the larger of the two being ω_+ . We can also check that the group velocity $d\omega/dk$ of both modes goes to zero at the zone boundary (Similarly the optical mode has zero group velocity at $k = 0$).

In Fig. 9.1 above, we have shown both modes at each value of k , such that we only need to show k within the first Brillouin zone. This is known as the *reduced zone scheme*. Another way to plot exactly the same dispersions is shown in Fig. 9.2 and is known as the *extended zone scheme*. Essentially you can think of this as “unfolding” the dispersions such that there is only one mode at each value of k . In this picture we have defined (for the first time) the *second Brillouin zone*.

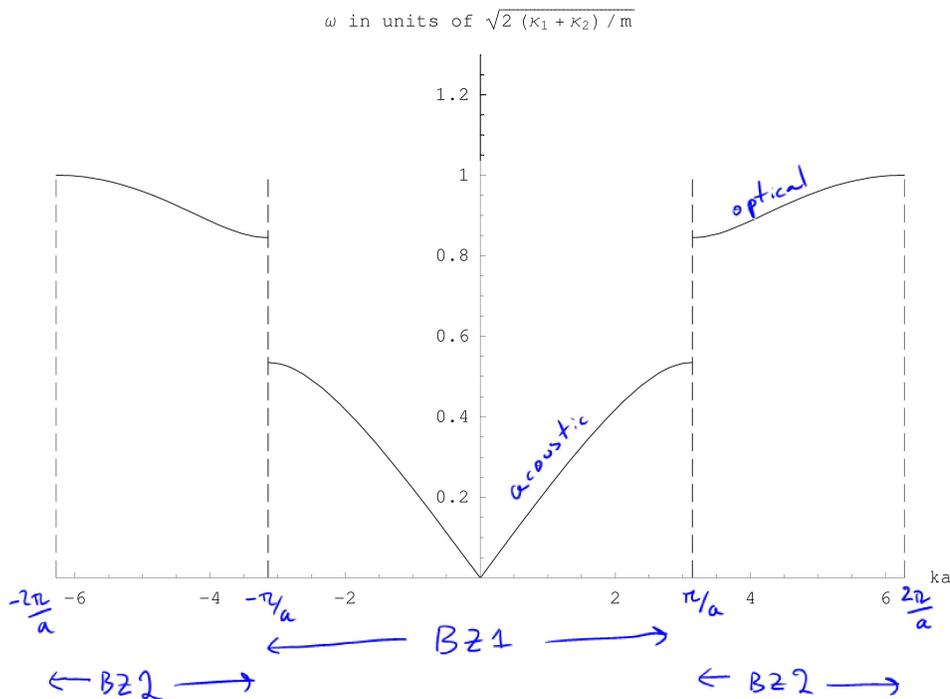


Figure 9.2: Dispersion Relation of Vibrations of the One Dimensional Diatomic Chain in the Extended Zone Scheme (Again choosing $\kappa_2 = 1.4\kappa_1$). Compare this to Fig. 9.1 above. The first Brillouin zone is labeled BZ1 and the second Brillouin zone is labeled BZ2.

Recall the first zone in 1d is defined as $|k| \leq \pi/a$. Analogously the second Brillouin zone is now $\pi/a \leq |k| \leq 2\pi/a$. In later chapters we will define the Brillouin zones more generally.

Here is an example where it is very useful to think using the extended zone scheme. We have been considering cases with $\kappa_2 > \kappa_1$, now let us consider what would happen if we take the limit of $\kappa_2 \rightarrow \kappa_1$. When the two spring constants become the same, then in fact the two atoms in the unit cell become identical, and we have a simple monatomic chain (which we discussed at length in the previous chapter). As such we should define a new smaller unit cell with lattice constant $a/2$, and the dispersion curve is now just a simple $|\sin|$ as it was in chapter 8.

Thus it is frequently useful if the two atoms in a unit cell are not too different from each

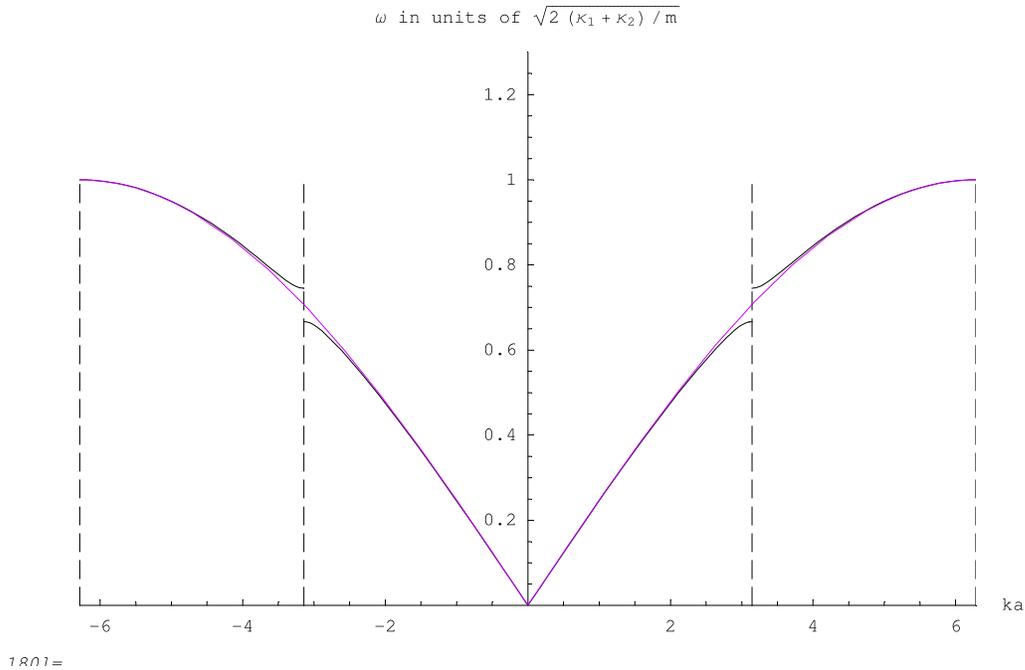


Figure 9.3: How a Diatomic Dispersion Becomes a Monatomic Dispersion When the Two Different Atoms Become the Same. (black) Dispersion relation of vibrations of the one dimensional diatomic chain in the extended zone scheme with κ_2 not too different from κ_1 . (blue) Dispersion relation when $\kappa_2 = \kappa_1$. In this case, the two atoms become exactly the same, and we have a monatomic chain with lattice spacing $a/2$. This single band dispersion precisely matches that calculated in chapter 8 above, only with the lattice constant redefined to $a/2$.

other, to think about the dispersion as being a small perturbation to a situation where all atoms are identical. When the atoms are made slightly different, a small gap opens up at the zone boundary, but the rest of the dispersion continues to look mostly as if it is the dispersion of the monatomic chain. This is illustrated in Fig. 9.3.

9.3 Summary of Vibrations of the One Dimensional Diatomic Chain

A number of key concepts are introduced in this chapter as well

- A unit cell is the repeated motif that comprises a crystal.
- The basis is the description of the unit cell with respect to a reference lattice.
- The lattice constant is the size of the unit cell (in 1d).
- If there are M atoms per unit cell we will find M normal modes at each wavevector k .

- One of these modes is an acoustic mode, meaning that it has linear dispersion at small k , whereas the remaining $M - 1$ are optical meaning they have finite frequency at $k = 0$.
- For the acoustic mode, all atoms in the unit cell move in-phase with each other, whereas for optical modes, they move out of phase with each other.
- Except for the acoustic mode, all other excitation branches have zero group velocity for $k = n\pi/a$ for any n .
- If all of the dispersion curves are plotted within the first Brillouin zone $|k| \leq \pi/a$ we call this the reduced zone scheme. If we “unfold” the curves such that there is only one excitation plotted per k , but we use more than one Brillouin zone, we call this the extended zone scheme.
- If the two atoms in the unit cell become identical, the new unit cell is half the size of the old unit cell. It is convenient to describe this limit in the extended zone scheme.

References

- Ashcroft and Mermin, chapter 22 (but not the 3d part)
- Ibach and Luth, section 4.3
- Kittel, chapter 4
- Hook and Hall, sections 2.3.2, 2.4, 2.5
- Burns, section 12.3

Chapter 10

Tight Binding Chain (Interlude and Preview)

In the previous two chapters we have considered the properties of vibrational waves (phonons) in a one dimensional system. At this point, we are going to make a bit of an excursion to consider electrons in solids again. The point of this excursion, besides being a preview of much of the physics that will re-occur later on, is to make the point that all waves in periodic environments (in crystals) are similar. In the previous two chapters we considered vibrational waves. In this chapter we will consider electron waves (Remember that in quantum mechanics particles are just as well considered to be waves!)

10.1 Tight Binding Model in One Dimension

We described the molecular orbital, or tight binding, picture for molecules previously in section 5.3.2. We also met the equivalent picture, or LCAO (linear combination of atomic orbitals) model of bonding for homework. What we will do here is consider a chain of such molecular orbitals to represent orbitals in a macroscopic (one dimensional) solid.

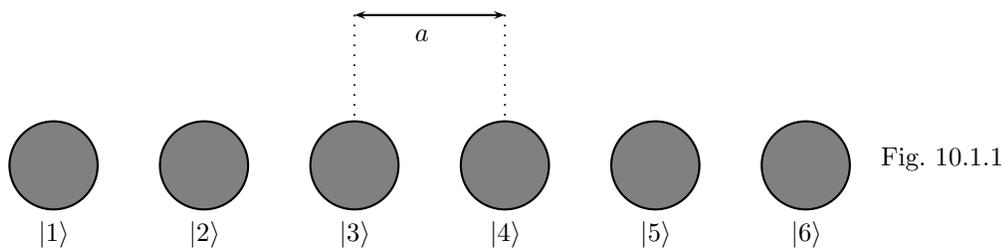


Fig. 10.1.1

In this picture, there is a single orbital on atom n which we call $|n\rangle$. For convenience we will assume that the system has periodic boundary conditions (i.e, there are N sites, and site N

is the same as site 0). Further we will assume that all of the orbitals are orthogonal to each other.

$$\langle n|m\rangle = \delta_{n,m} \quad (10.1)$$

Let us now take a general trial wavefunction of the form

$$|\Psi\rangle = \sum_n \phi_n |n\rangle$$

As we showed for homework, the effective Schrödinger equation for this type of tight-binding model can be written as

$$\sum_m H_{nm} \phi_m = E \phi_n \quad (10.2)$$

where H_{nm} is the matrix element of the Hamiltonian

$$H_{nm} = \langle n|H|m\rangle$$

As mentioned previously when we studied the molecular orbital model, this Schrödinger equation is actually a variational approximation. For example, instead of finding the exact ground state, it finds the best possible ground state made up of the orbitals that we have put in the model.

One can make the variational approach increasingly better by expanding the Hilbert space and putting more orbitals into the model. For example, instead of having only one orbital $|n\rangle$ at a given site, one could consider many $|n, \alpha\rangle$ where α runs from 1 to some number p . As p is increased the approach becomes increasingly more accurate and eventually is essentially exact. This method of using tight-binding like orbitals to increasingly well approximate the exact Schrödinger equation is known as LCAO (linear combination of atomic orbitals). However, one complication (which we treat only in one of the additional homework assignments) is that when we add many more orbitals we typically have to give up our nice orthogonality assumption, i.e., $\langle n, \alpha|m, \beta\rangle = \delta_{nm}\delta_{\alpha\beta}$ no longer holds. This makes the effective Schrödinger equation a bit more complicated, but not fundamentally different. (See comments in section 5.3.2 above).

At any rate, in the current chapter we will work with only one orbital per site and we assume the orthogonality Eq. 10.1.

We write the Hamiltonian as

$$H = K + \sum_j V_j$$

where $K = p^2/(2m)$ is the kinetic energy and V_j is the Coulomb interaction of the electron with the nucleus at site j ,

$$V_j = V(r - r_j)$$

where r_j is the position of the j^{th} nucleus.

With these definitions we have

$$H|m\rangle = (K + V_m)|m\rangle + \sum_{j \neq m} V_j|m\rangle$$

Now, we should recognize that $K + V_m$ is the Hamiltonian which we would have if there were only a single nucleus (the m^{th} nucleus) and no other nuclei in the system. Thus, if we take the tight-binding orbitals $|m\rangle$ to be the atomic orbitals, then we have

$$(K + V_m)|m\rangle = \epsilon_{\text{atomic}}|m\rangle$$

where ϵ_{atomic} is the energy of an electron on nucleus m in the absence of any other nuclei.

Thus we can write

$$H_{n,m} = \langle n|H|m\rangle = \epsilon_{atomic} \delta_{n,m} + \sum_{j \neq m} \langle n|V_j|m\rangle$$

We now have to figure out what the final term of this equation is. The meaning of this term is that, via the interaction with some nucleus which is not the m^{th} , an electron on the m^{th} atom can be transferred to the n^{th} . Generally this can only happen if n and m are very close to each other. Thus, we write

$$\sum_{j \neq m} \langle n|V_j|m\rangle = \begin{cases} V_0 & n = m \\ -t & n = m \pm 1 \\ 0 & \text{otherwise} \end{cases} \quad (10.3)$$

which defines both V_0 and t . (The V_0 term here does not hop an electron from one site to another, but rather just shifts the energy on a given site.) Note by translational invariance of the system, we expect that the result should depend only on $n - m$, which this form does. These two types of terms V_0 and t are entirely analogous to two types of terms V_{cross} and t that we met in section 5.3.2 above when we studied covalent bonding of two atoms¹. The situation here is similar except that now there are many nuclei instead of just two.

With the above matrix elements we obtain

$$H_{n,m} = \epsilon_0 \delta_{n,m} - t (\delta_{n+1,m} + \delta_{n-1,m}) \quad (10.4)$$

where we have now defined²

$$\epsilon_0 = \epsilon_{atomic} + V_0$$

This Hamiltonian is a very heavily studied model, known as the tight binding chain. Here t is known as the hopping term, as it allows the Hamiltonian (which generates time evolution) to move the electron from one site to another, and it has dimensions of energy. It stands to reason that the magnitude of t depends on how close together the orbitals are — becoming large when the orbitals are close together and decaying exponentially when they are far apart.

10.2 Solution of the Tight Binding Chain

The solution of the tight binding model in one dimension (the tight binding chain) is very analogous to what we did to study vibrations (and hence the point of presenting the tight binding model at this point!). We propose an ansatz solution

$$\phi_n = \frac{e^{-ikna}}{\sqrt{N}} \quad (10.5)$$

where the denominator is included for normalization where there are N sites in the system. We now plug this ansatz into the Schrödinger equation Eq. 10.2. Note that in this case, there is no frequency in the exponent of our ansatz. This is simply because we are trying to solve the time-independent Schrödinger equation. Had we used the time dependent equation, we would need a factor of $e^{i\omega t}$ as well!

¹Just to be confusing, atomic physicists sometimes use J where I have used t here.

²Once again ϵ_0 is *not* a dielectric constant or the permittivity of free space, but rather just the energy of having an electron sit on a site.

As with vibrations, it is obvious that $k \rightarrow k + 2\pi/a$ gives the same solution. Further, if we consider the system to have periodic boundary conditions with N sites (length $L = Na$), the allowed values of k are quantized in units of $2\pi/L$. As with Eq. 8.6 there are precisely N possible different solutions of the form of Eq. 10.5.

Plugging the ansatz into the left side of the Schrödinger equation 10.2 and then using Eq. 10.4 gives us

$$\sum_m H_{n,m} \phi_m = \epsilon_0 \frac{e^{-ikna}}{\sqrt{N}} - t \left(\frac{e^{-ik(n+1)a}}{\sqrt{N}} + \frac{e^{-ik(n-1)a}}{\sqrt{N}} \right)$$

which we set equal to the right side of the Schrödinger equation

$$E \phi_n = E \frac{e^{-ikna}}{\sqrt{N}}$$

To obtain the spectrum

$$E = \epsilon_0 - 2t \cos(ka) \quad (10.6)$$

which looks rather similar to the *phonon* spectrum of the one dimensional monatomic chain which was (See Eq. 8.2)

$$\omega^2 = 2 \frac{\kappa}{m} - 2 \frac{\kappa}{m} \cos(ka)$$

Note however, that in the electronic case one obtains the energy whereas in the phonon case one obtains the *square* of the frequency.

This dispersion curve is shown in Fig. 10.1. Analogous to the phonon case, it is periodic in $k \rightarrow k + 2\pi/a$. Further, analogous to the phonon case, the dispersion always has zero group velocity (is flat) for $k = n\pi/a$ for n any integer (i.e., at the Brillouin zone boundary).

Note that unlike free electrons, the electron dispersion here has a maximum energy as well as a minimum energy. Electrons only have eigenstates within a certain energy *band*. The word “band” is used both to describe the energy range for which eigenstates exist, as well as to describe one connected branch of the dispersion curve (In this picture there is only a single mode at each k , hence one branch, hence a single band).

The energy difference from the bottom of the band to the top is known as the *bandwidth*. Within this bandwidth (between the top and bottom of the band) for any energy there exists (at least one) k state having that energy. For energies outside of the bandwidth there are no k -states with that energy.

The bandwidth (which in this model is $4t$) is determined by the magnitude of the hopping, which, as mentioned above, depends on the distance between nuclei³. As a function of the interatomic spacing then the bandwidth increases as shown in Fig 10.2. On the right of this diagram there are N states, each one being an atomic orbital $|n\rangle$. On the left of the diagram these N states form a band, yet as discussed above, there remain precisely N states. (This should not surprise us, being that we have not changed the dimension of the Hilbert state, we have just expressed it in terms of the complete set of eigenvalues of the Hamiltonian). Note that the average energy of a state in this band remains always zero.

Aside: Note that if the band is not completely filled, the *total* energy of all of the electrons *decreases* as the atoms are moved together and the band width increases. (Since the average energy remains zero, but some of the higher energy states are not filled). This decrease in energy is precisely the binding force of a “metallic

³Since the hopping t depends on an overlap between orbitals on adjacent atoms (See Eq. 10.3), in the limit that the atoms are well separated, the bandwidth will increase exponentially as the atoms are pushed closer together.

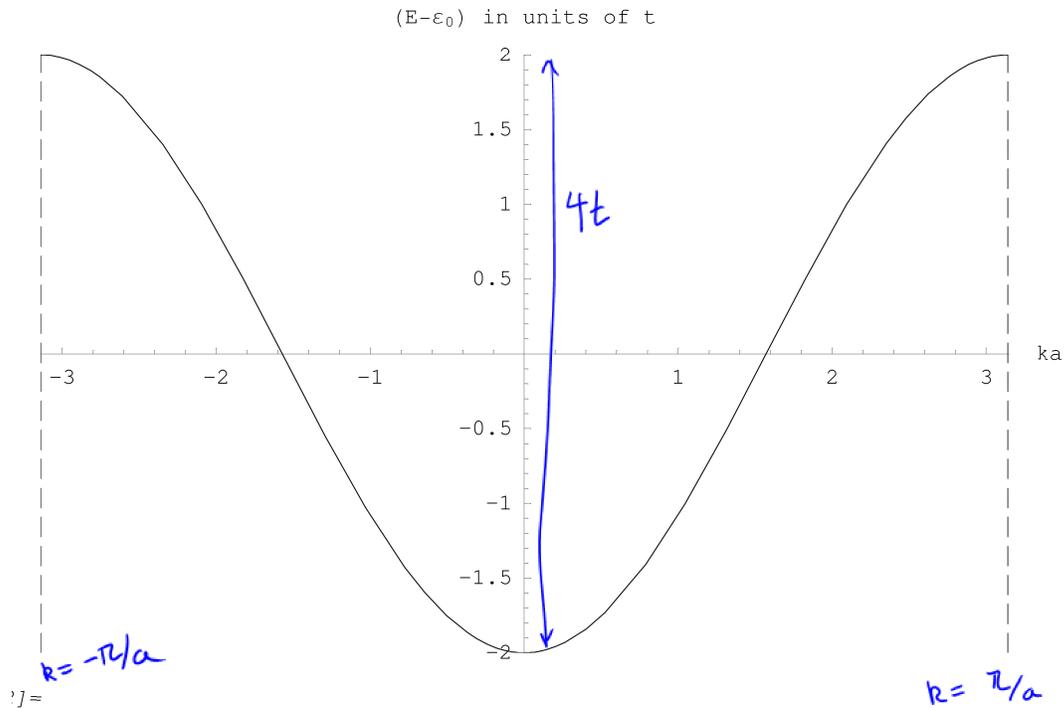


Figure 10.1: Dispersion of the Tight Binding Chain.

bond" which we discussed in section 5.5.⁴ We also mentioned previously that one property of metals is that they are typically soft and malleable. This is a result of the fact that the electrons that hold the atoms together are mobile — in essence, because they are mobile, they can readjust their positions somewhat as the crystal is deformed.

Near the bottom of the band, the dispersion is parabolic. For our above dispersion (Eq. 10.6), expanding for small k , we obtain

$$E(k) = \text{Constant} + ta^2k^2$$

[Note that for $t < 0$, the energy minimum is at the Brillouin zone boundary $k = \pi/a$. In this case we would expand for k close to π/a instead of for k close to 0]. The resulting parabolic behavior is similar to that of free electrons which have a dispersion

$$E_{free}(k) = \frac{\hbar^2 k^2}{2m}$$

We can therefore view the bottom of the band as being almost like free electrons, except that we have to define a new *effective mass* which we call m^* such that

$$\frac{\hbar^2 k^2}{2m^*} = ta^2k^2$$

⁴Of course we have not considered the repulsive force between neighboring nuclei, so the nuclei do not get too close together. As in the case of the covalent bond considered above in section 5.3.2, some of the Coulomb repulsion between nuclei will be canceled by V_{cross} (here V_0) the attraction of the electron on a given site to other nuclei.

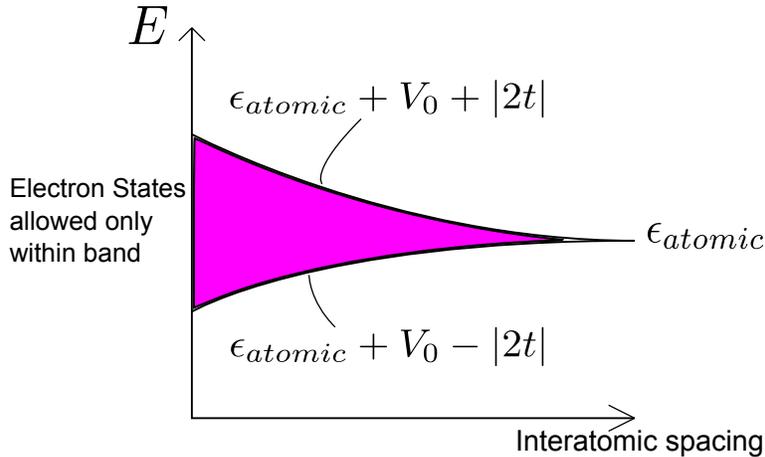


Figure 10.2: Caricature of the Dependence of Bandwidth on Interatomic Spacing.

which gives us

$$m^* = \frac{\hbar^2}{2ta^2}$$

In other words, the *effective mass* m^* is defined such that the dispersion of the bottom of the band is exactly like the dispersion of free particles of mass m^* . (We will discuss effective mass in much more depth in chapter 16 below. This is just a quick first look at it.) Note that this mass has nothing to do with the actual mass of the electron, but rather depends on the hopping matrix element t . Further we should keep in mind that the k that enters into the dispersion relationship is actually the crystal momentum, not the actual momentum of the electron (recall that crystal momentum is defined only modulo $2\pi/a$). However, so long as we stay at very small k , then there is no need to worry about the periodicity of k which occurs. Nonetheless, we should keep in mind that if electrons scatter off of other electrons, or off of phonons, it is crystal momentum that is conserved. (See the discussion in section 8.4).

10.3 Introduction to Electrons Filling Bands

We now imagine that our tight binding model is actually made up of atoms and each atom “donates” one electron into the band (i.e., the atom has *valence* one). Since there are N possible k -states in the band, and electrons are fermions, you might guess that this would precisely fill the band. However, there are two possible spin states for an electron at each k , so in fact, this then only half-fills the band. This is depicted in the left of Fig. 10.3. The filled states (shaded) in this picture are filled with both up and down spins.

It is crucial in this picture that there is a Fermi surface — the points where the shaded meets the unshaded region. If a small electric field is applied to the system, it only costs a very small amount of energy to shift the Fermi surface as shown in the right of Fig. 10.3, populating a few k -states moving right and de-populating some k -states moving left. In other words, the state of the system responds by changing a small bit and a current is induced. As such, this system is a *metal* in that it conducts electricity. Indeed, crystals of atoms that are mono-valent are very frequently metals!

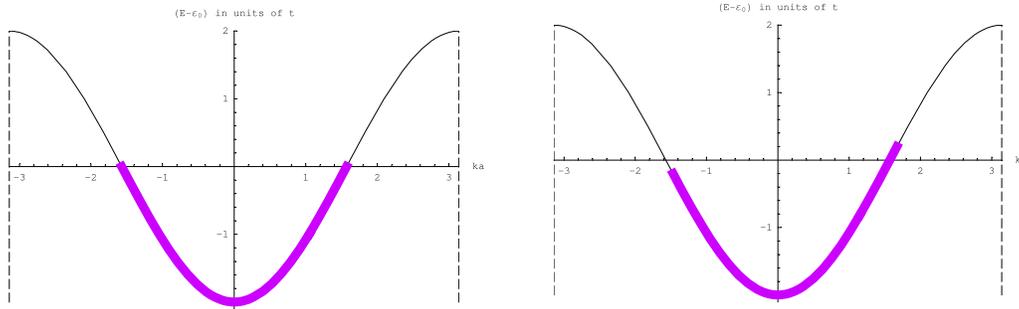


Figure 10.3: Left: If each atom has valence 1, then the band is half-filled. The states that are shaded are filled with both up and down spin electrons. The Fermi surface is the boundary between the filled and unfilled states. Right: When a small electric field is applied, at only a small cost of energy, the Fermi sea can shift slightly thus allowing current to run.

On the other hand, if each atom in our model were di-valent (donates two electrons to the band) then the band would be entirely full of electrons. In fact, it does not matter if we think about this as being a full band where every k -state $|k\rangle$ is filled with two electrons (one up and one down), or a filled band where every site $|n\rangle$ is filled — these two statements describe the same multi-electron wavefunction. In fact, there is a single unique wavefunction that describes this completely filled band.

In the case of the filled band, were one to apply a small electric field to this system, the system cannot respond at all. There is simply no freedom to repopulate the occupation of k -states because every state is already filled. We conclude an important principle,

Principle: A filled band carries no current.

Thus our example of a di-valent tightbinding model is an insulator. (This type of insulator is known as a *band insulator*). Indeed, many systems of di-valent atoms are insulators (although in a moment we will discuss how di-valent atoms can also form metals).

10.4 Multiple Bands

In the above model, we considered only the case where there is a single atom in the unit cell and a single orbital per atom. However, more generally we might consider a case where we have multiple orbitals per unit cell.

One possibility is to consider one atom per unit cell, but several orbitals per atom⁵. Analogous to what we found with the above tight binding model, when the atoms are very far apart, one has only the atomic orbitals on each atom. However, as the atoms are moved closer together, the orbitals merge together and the energies spread to form bands⁶. Analogous to Fig. 10.2 we

⁵Each atom actually has an infinite number of orbitals to be considered. But only a small number of them are filled, and within our level of approximation, we can only consider very few of them.

⁶This picture of atomic orbitals in the weak hopping limit merging together to form bands does not depend on the fact that the crystal of atoms is ordered. Glasses and amorphous solids can have this sort of band structure as well!

have shown how this occurs for the two band case in Fig. 10.4.

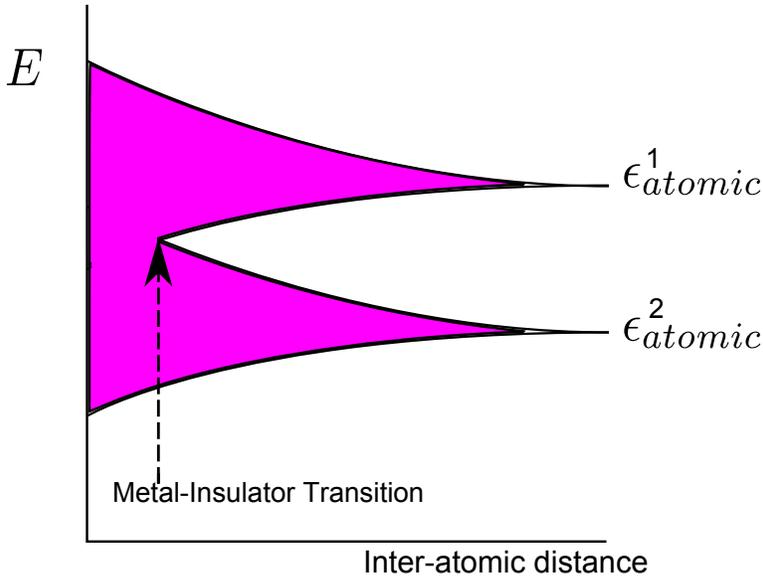


Figure 10.4: Caricature of Bands for a Two-Band Model as a Function of Interatomic Spacing . In the atomic limit, the orbitals have energies ϵ_{atomic}^1 and ϵ_{atomic}^2 . If the system has valence one (per unit cell), then in the atomic limit, the lower orbital is filled and the upper orbital is empty. When the atoms are pushed together, the lower band will remain filled, and the upper will remain empty, until the bands start to overlap, whereupon we may have two bands both partially filled, which becomes a metal.

A very similar situation occurs when we have two atoms per unit cell but only one orbital per atom. We will do a problem like this for homework⁷. However, the general result will be quite analogous to what we found for vibrations of a diatomic chain in chapter 9.

In Fig. 10.5 we show the spectrum of a tight-binding model with two different atoms per unit cell – each having a single orbital. We have shown results here in both the reduced and extended zone schemes.

As for the case of vibrations, we see that there are now two possible energy eigenstates at each value of k . In the language of electrons, we say that there are two bands (we do not use the words “acoustic” and “optical” for electrons, but the idea is similar). Note that there is a gap between the two bands where there are simply no energy eigenstates.

Let us think for a second about what should result in this situation. If each atom (of either type) were divalent, then the two electrons donated would completely fill the single orbital on each site. In this case, both bands would be completely filled with both spin-up and spin-down electrons.

On the other hand, if each atom (of either type) is monovalent, then this means exactly half of the states of the system should be filled. However, here, when one fills half of the states of the system, then all of the states of the lower band are completely filled (with both spins) but all of

⁷The homework problem is sufficiently simplified that the bands do not overlap as they do in figure 10.4. One can obtain overlapping bands by including second-neighbor hopping as well as neighbor hopping. (If you are brave you might try it!).

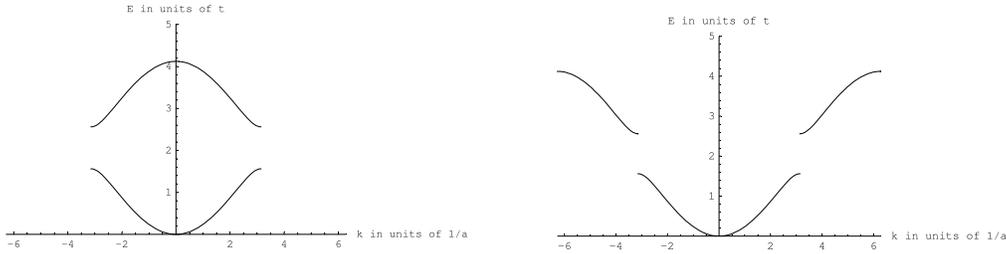


Figure 10.5: Diatomic Tight Binding Dispersion in One Dimension. Left: Reduced Zone scheme. Right: Extended Zone scheme.

the states in the upper band are completely empty. In the extended zone scheme it appears that a gap has opened up precisely where the Fermi surface is! (at the Brillouin zone boundary!)

In the situation where a lower band is completely filled but an upper band is completely empty, if we apply a weak electric field to the system can current flow? In this case, one cannot rearrange electrons within the lower band, but one can remove an electron from the lower band and put it in the upper band in order to change the overall (crystal) momentum of the system. However, moving an electron from the lower band requires a finite amount of energy — one must overcome the gap between the bands. As a result, for small enough electric fields (and at low temperature), this cannot happen. We conclude that a filled band is an insulator as long as there is a finite gap to any higher empty bands.

As with the single band case, one can imagine the magnitude of hopping changing as one changes the distance between atoms. When the atoms are far apart, then one is in the atomic limit, but these atomic states spread into bands as the atoms get closer together as shown in Fig. 10.4.

For the case where each atom is mono-valent, in the atomic limit, half of the states are filled — that is the lower energy atomic orbital is filled with both spin-up and spin down electrons whereas the higher energy orbital is completely empty. (I.e., an electron is transferred from the higher energy atom to the lower energy atom and this completely fills the lower energy band). As the atoms are brought closer together, the atomic orbitals spread into bands (the hopping t increases). However, at some point the bands get so wide that their energies overlap⁸ — in which case there is no gap to transfer electrons between bands, and the system becomes a metal as marked in Fig. 10.4. (If it is not clear how bands may overlap, consider, for example the right side of Fig. 15.2. Band overlaps may occur — in fact, they often occur! — when we consider systems that are two and three dimensional.)

10.5 Summary of Tight Binding Chain

- Solving tight-binding Shroedinger equation for electron waves is very similar to solving equations for vibrational (phonon) waves. The structure of the reciprocal lattice and the Brillouin

⁸As mentioned above, in our simplified model one needs to consider second neighbor hopping to get overlapping bands.

zone remains the same.

- Obtain energy bands where energy eigenstates exist, and gaps between bands.
- Zero hopping is the atomic limit, as hopping increases, atomic orbitals spread into bands.
- Energies are parabolic in k near bottom of band — looks like free electrons, but with a modified effective mass.
- A filled band with a gap to the next band is an insulator (a *band* insulator), a partially filled band has a Fermi surface and is a metal.
- Whether a band is filled depends on the valence of the atoms.
- As we found for phonons, gaps open at Brillouin zone boundaries. Group velocities are also zero at zone boundaries.

References

No book has an approach to tight binding that is exactly like what we have here. The books that come closest do essentially the same thing, but in three dimensions (which complicates life a bit). These books are:

- Ibach and Luth, section 7.3
- Kittel, chapter 9, section on tight-binding
- Burns, section 10.9, and 10.10.
- Singleton, chapter 4.

Possibly the nicest (albeit short) description is given by

- Dove, section 5.5.5

Also a nice short description of the physics (without any detail is given by)

- Rosenberg, section 8.19.

Finally, an alternative approach to tight binding is given by

- Hook and Hall, section 4.3.

The discussion of Hook and Hall is good (and they consider one dimension, which is nice), but they insist on using time dependent Schrödinger equation, which is annoying.

Part IV

Geometry of Solids

Chapter 11

Crystal Structure

Having introduced a number of important ideas in one dimension, we must now deal with the fact that our world is actually spatially three dimensional. While this adds a bit of complication, really the important concepts are no harder in three dimensions than they were in one dimension. Some of the most important ideas we have already met in one dimension, but we will reintroduce more generally here.

There are two parts that might be difficult here. First, we do need to wrestle with a bit of geometry. Hopefully most will not find this too hard. Secondly will also need to establish a language in order to describe structures in two and three dimensions intelligently. As such, much of this chapter is just a list of definitions to be learned, but unfortunately this is necessary in order to be able to carry further at this point.

11.1 Lattices and Unit Cells

Definition 11.1.1. A *Lattice*¹ is an infinite set of points defined by integer sums of a set of linearly independent primitive lattice² vectors

For example, in two dimensions, as shown in figure 11.1 the lattice points are described as

$$\mathbf{R}_{[n_1 n_2]} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 \quad n_1, n_2 \in \mathbb{Z} \quad (2d)$$

with \mathbf{a}_1 and \mathbf{a}_2 being the primitive lattice vectors and n_1 and n_2 being integers. In three dimensions points of a lattice are analogously indexed by three integers

$$\mathbf{R}_{[n_1 n_2 n_3]} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3 \quad n_1, n_2, n_3 \in \mathbb{Z} \quad (3d) \quad (11.1)$$

¹Warning: Some books (Ashcroft and Mermin in particular) refer to this as a *Bravais Lattice*. This enables them to use the term *Lattice* to describe other things that we would not call a lattice (cf, the honeycomb). However, the definition we use here is more common, and more correct mathematically as well. [Thank you, Mike Glazer, for catching this].

²Very frequently “primitive lattice vectors” are called “primitive basis vectors”, although the former is probably more precise. Furthermore, we have already used the word “basis” before in chapter 9.1, and unfortunately, here this is a different use of the same word. At any rate, we will try to use “primitive lattice vector” to avoid such confusion.

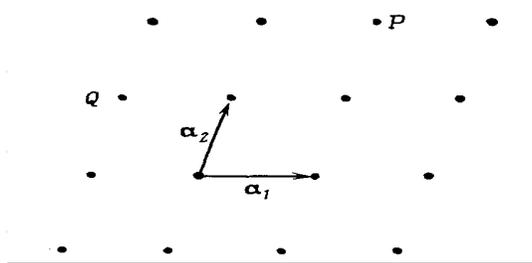


Figure 11.1: A lattice is defined as integer sums of a set of primitive lattice vectors.

Note that in one dimension this definition of a lattice fits with our previous description of a lattice as being the points $R = na$ with n an integer.

It is important to point out that in two and three dimensions, the choice of primitive lattice vectors is not unique³ as show in figure 11.2. (In 1d, the single primitive lattice vector is unique up to the sign (direction) of a).

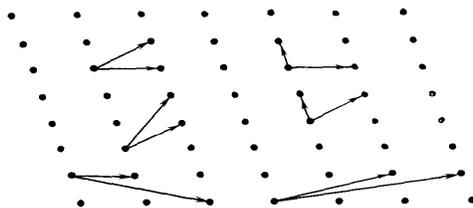


Figure 11.2: The choice of primitive lattice vectors for a lattice is not unique.

It turns out that there are several definitions that are entirely equivalent to the one we have just given:

Equivalent Definition 11.1.1.1. A *Lattice* is an infinite set of vectors where addition of any two vectors in the set gives a third vector in the set.

It is easy to see that our above first definition 11.1.1 implies the second one 11.1.1.1. Here is a less crisply defined, but sometimes more useful definition.

Equivalent Definition 11.1.1.2. A *Lattice* is a set of points where the environment of any given point is equivalent to the environment of any other given point.

³Given a set of primitive lattice vectors \mathbf{a}_j a new set of primitive lattice vectors may be constructed as $\mathbf{b}_i = \sum_j m_{ij} \mathbf{a}_j$ so long as m_{ij} is an invertible matrix with integer entries and the inverse matrix $[m^{-1}]_{ij}$ also has integer entries.

It turns out that any periodic structure can be expressed as a lattice of repeating motifs. A cartoon of this statement is shown in Fig. 11.3.

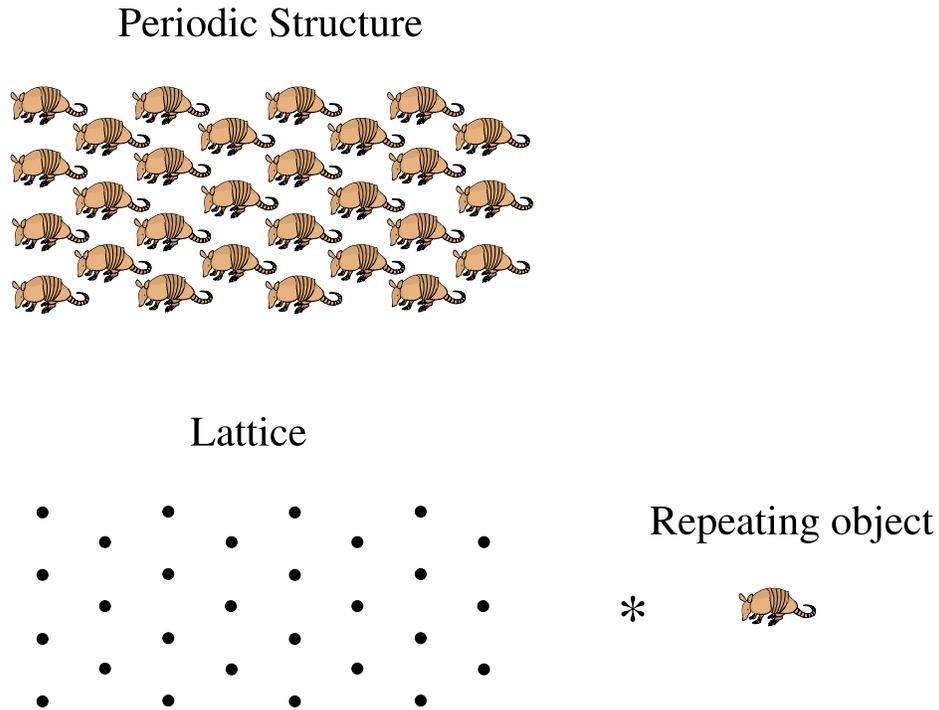


Figure 11.3: Any periodic structure can be represented as a lattice of repeating motifs.

One should be cautious however, that not all periodic arrangements of points are lattices. The honeycomb⁴ shown in Fig. 11.4 is *not* a lattice. This is obvious from the third definition 11.1.1.2: The environment of point P and point R are actually different — point P has a neighbor directly above it (the point R), whereas point R has no neighbor directly above.

In order to describe a honeycomb (or other more complicated arrangements of points) we have the idea of a unit cell, which we have met before in section 9.1 above. Generally we have

Definition 11.1.2. A *unit cell* is a region of space such that when many identical units are stacked together it tiles (completely fills) all of space and reconstructs the full structure

An equivalent (but less rigorous) definition is

Equivalent Definition 11.1.2.1. A *unit cell* is the repeated motif which is the elementary building block of the periodic structure.

To be more specific we frequently want to work with the smallest possible unit cell

⁴One should be careful *not* to call this a hexagonal lattice. First of all, by our definition it, is not a lattice at all since all points do not have the same environment. Secondly, some people use the term “hexagonal” to mean what the rest of us call a triangular lattice: a lattice of triangles where each point has six nearest neighbor points. (See Fig 11.6 below)

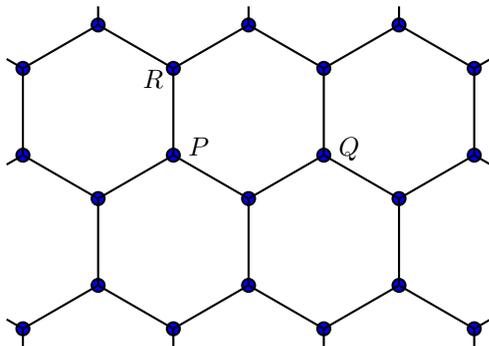


Figure 11.4: The honeycomb is not a lattice. Points P and R are inequivalent. (Points P and Q are equivalent)

Definition 11.1.3. A *primitive unit cell* for a periodic crystal is a unit cell containing only a single lattice point.

As mentioned above in section 9.1 the definition of the unit cell is never unique. This is shown, for example, in Fig. 11.5

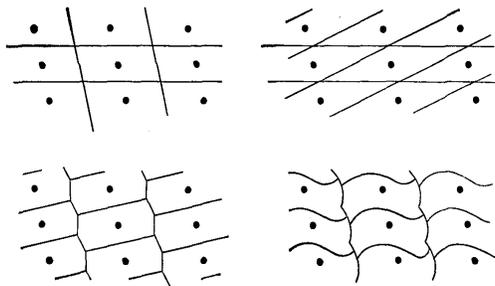


Figure 11.5: The choice of a unit cell is not unique. All of these unit cells reconstruct the same crystal.

Sometimes it is useful to define a unit cell which is not primitive in order to have it simpler to work with. This is known as a *conventional unit cell*. Almost always these conventional unit cells are chosen so as to have orthogonal axes.

Some examples of possible unit cells are shown for the triangular lattice in Fig. 11.6. In this

figure the conventional unit cell (upper left) is chosen to have orthogonal axes — which is often easier to work with than axes which are non-orthogonal.

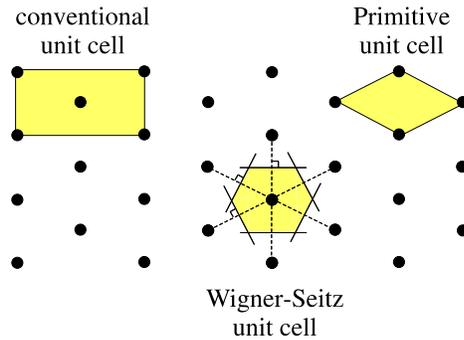


Figure 11.6: Some unit cells for the triangular lattice.

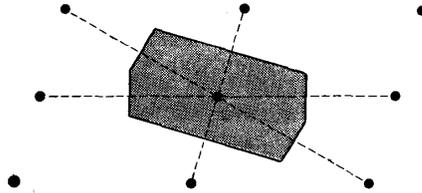


Figure 11.7: The Wigner-Seitz construction for a lattice in 2d.

A note about counting the number of lattice points in the unit cell: It is frequently the case that we will work with unit cells where the lattice points live at the corners (or edges) of the cells. When a lattice point is on the boundary of the unit cell, it should only be counted fractionally depending on what fraction of the point is actually in the cell. So for example in the conventional unit cell shown in Fig. 11.6, there are two lattice points within this cell. Obviously there is one point in the center, then four points at the corners — each of which is one quarter inside the cell, so we obtain $2 = 1 + 4(\frac{1}{4})$ points in the cell. (Since there are two points in this cell, it is by definition, not primitive. Similarly for the primitive cell shown in this figure (upper right), the two lattice points at the left and the right have a 60° degree slice (which is $1/6$ of a circle) inside the cell. The two points at the top and the bottom have $1/3$ of the point inside the unit cell. Thus this unit cell contains $2(\frac{1}{3}) + 2(\frac{1}{6}) = 1$ point, and is thus primitive. Note however, that we can just imagine shifting the unit cell a tiny amount in almost any direction such that a single lattice point is completely inside the unit cell and the others are completely outside the unit cell. This sometimes makes counting much easier.

Also shown in Fig. 11.6 is a so-called *Wigner-Seitz unit cell*⁵.

⁵Eugene Wigner was yet another Nobel laureate who was another one of the truly great minds of the last century

Definition 11.1.4. Given a lattice point, the set of all points in space which are closer to that given lattice point than to any other lattice point constitute the *Wigner-Seitz cell* of the given lattice point.

There is a rather simple scheme for constructing such a Wigner-Seitz cell: choose a lattice point and draw lines to all of its possible near neighbors (not just its nearest neighbors). Then draw perpendicular bisectors of all of these lines. The perpendicular bisectors bound the Wigner-Seitz cell⁶. It is always true that the Wigner-Seitz construction for a lattice gives a primitive unit cell. In figure 11.7 we show another example of the Wigner-Seitz construction for a two dimensional lattice. A similar construction can be performed in three dimensions in which case one must construct perpendicular-bisecting planes to bound the Wigner-Seitz cell.

The description of objects in the unit cell in terms of the reference point in the unit cell is known as a “basis”. (This is the same definition of “basis” that we used in section 9.1 above).

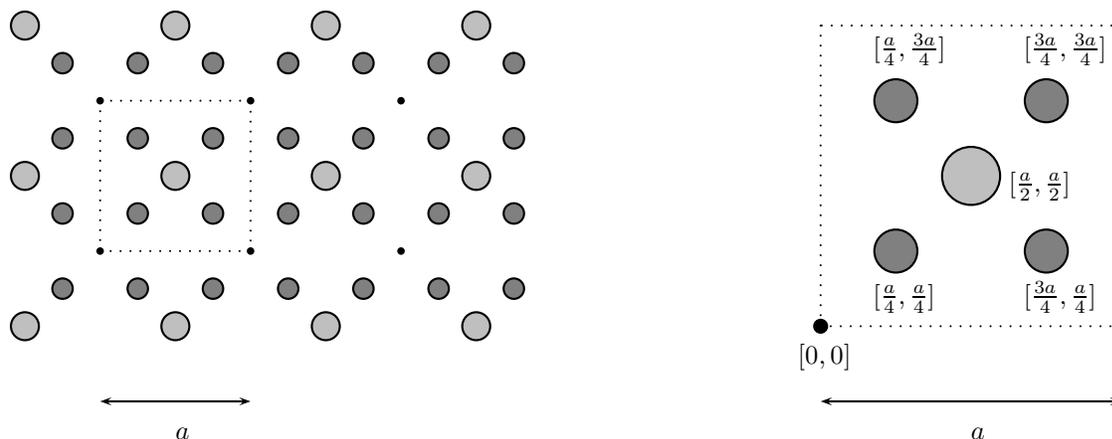


Figure 11.8: Left: A periodic structure in two dimensions. A unit cell is marked with the dotted lines. Right: A blow-up of the unit cell with the coordinates of the particles in the unit cell with respect to the reference point in the lower left hand corner. The basis is the description of the atoms along with these positions.

In Fig. 11.8 we show a periodic structure in two dimension made of two types of atoms. On the right we show a primitive unit cell (expanded) with the position of the atoms given with respect to the reference point of the unit cell which is taken to be the lower left-hand corner. We can describe the basis of this crystal as follows:

of physics. Fredrick Seitz was far less famous, but gained notoriety in his later years by being a consultant for the tobacco industry, a strong proponent of the Regan-era Star-Wars missile defense system, and a prominent sceptic of global warming. He passed away in 2007.

⁶This Wigner-Seitz construction can be done on an irregular collection of points as well as on a periodic lattice. For such an irregular set of point the resulting construction is known as a Voronoi cell.

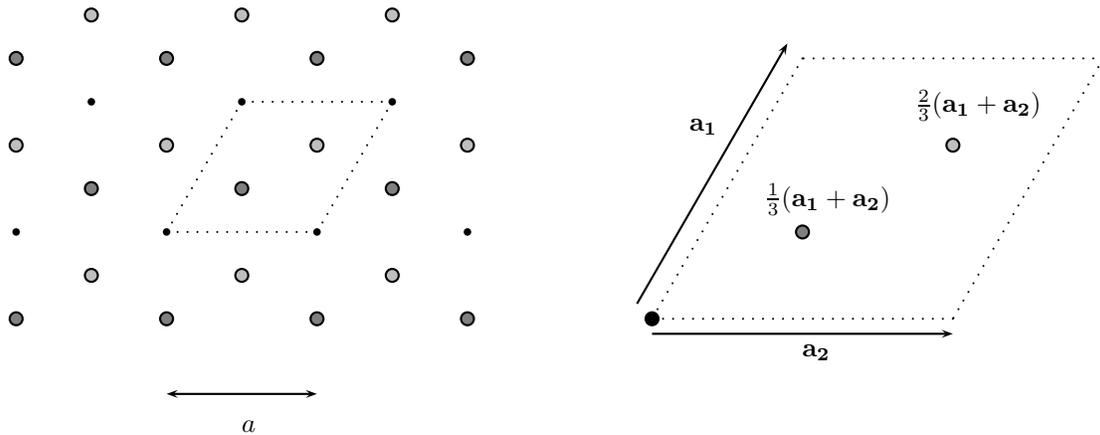


Figure 11.9: The honeycomb from Fig. 11.4 with the two inequivalent points of the unit cell given different shades. The unit cell is outlined dotted on the left and the corners of the unit cell are marked with small black dots. On the right the unit cell is expanded and coordinates are given with respect to the reference point written.

Basis for crystal in Fig. 11.8 =	
Large Light Gray Atom	Position= $[a/2, a/2]$
Small Dark Gray Atoms	Position= $[a/4, a/4]$
	$[a/4, 3a/4]$
	$[3a/4, a/4]$
	$[3a/4, 3a/4]$

The reference points forming the square lattice have positions

$$\mathbf{R}_{[n_1 n_2]} = [a n_1, a n_2] = a n_1 \hat{x} + a n_2 \hat{y} \quad (11.2)$$

with n_1, n_2 integers so that the large light gray atoms have positions

$$\mathbf{R}_{[n_1 n_2]}^{\text{light-gray}} = [a n_1, a n_2] + [a/2, a/2]$$

whereas the small dark gray atoms have positions

$$\begin{aligned} \mathbf{R}_{[n_1 n_2]}^{\text{dark-gray1}} &= [a n_1, a n_2] + [a/4, a/4] \\ \mathbf{R}_{[n_1 n_2]}^{\text{dark-gray2}} &= [a n_1, a n_2] + [a/4, 3a/4] \\ \mathbf{R}_{[n_1 n_2]}^{\text{dark-gray3}} &= [a n_1, a n_2] + [3a/4, a/4] \\ \mathbf{R}_{[n_1 n_2]}^{\text{dark-gray4}} &= [a n_1, a n_2] + [3a/4, 3a/4] \end{aligned}$$

In this way you can say that the positions of the atoms in the crystal are “the lattice plus the basis”.

We can now return to the case of the honeycomb shown in Fig. 11.4 above. The same honeycomb is shown in Fig. 11.9 as well with the lattice and the basis explicitly shown. Here, the reference points (small black dots) form a (triangular) lattice, where we can write the primitive lattice vectors as

$$\begin{aligned}\mathbf{a}_1 &= a \hat{x} \\ \mathbf{a}_2 &= (a/2) \hat{x} + (a\sqrt{3}/2) \hat{y}\end{aligned}$$

In terms of the reference points of the lattice, the basis for the primitive unit cell, i.e., the coordinates of the two larger circles with respect to the reference point, are given by $\frac{1}{3}(\mathbf{a}_1 + \mathbf{a}_2)$ and $\frac{2}{3}(\mathbf{a}_1 + \mathbf{a}_2)$.

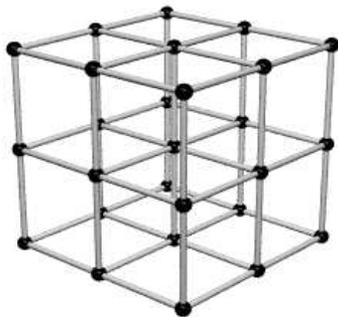


Figure 11.10: A simple cubic lattice

11.2 Lattices in Three Dimensions

The simplest lattice in three dimensions is the simple cubic lattice shown in Fig. 11.10 (sometimes known as cubic “P” or cubic-primitive lattice). The primitive unit cell in this case can most conveniently be taken to be single cube — which includes 1/8 of each of its eight corners.

In fact, real crystals of atoms are rarely simple cubic⁷. To understand why this is so, think of an atom as a small sphere. When you assemble spheres into a simple cubic lattice you find that it is a very inefficient way to pack the spheres together — in that you are left with a lot of empty space in the center of the unit cells, and this turns out to be energetically unfavorable in most cases.

Only slightly more complicated than the simple cubic lattice are the *tetragonal* and *orthorhombic* lattices where the axes remain perpendicular, but the primitive lattice vectors may be of different lengths (shown in Fig 11.11). The orthorhombic unit cell has three different lengths of its perpendicular primitive lattice vectors, whereas the tetragonal unit cell has two lengths the same and one different.

Conventionally to represent a given vector amongst the infinite number of possible lattice vectors in a lattice, one writes

$$[uvw] = u\mathbf{a}_1 + v\mathbf{a}_2 + w\mathbf{a}_3 \quad (11.3)$$

⁷Of all of the chemical elements, Polonium is the only one which forms a simple cubic lattice.

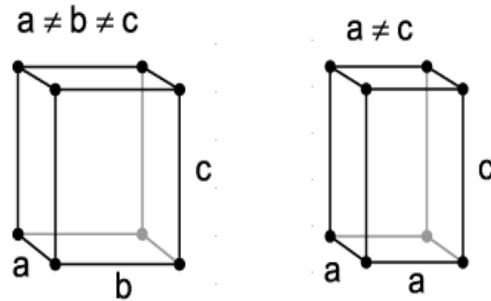


Figure 11.11: Unit cells for orthorhombic (left) and tetragonal (right) lattices.

where u, v and w are integers. For cases where the lattice vectors are orthogonal, the basis vectors \mathbf{a}_1 , \mathbf{a}_2 , and \mathbf{a}_3 are assumed to be in the \hat{x} , \hat{y} and \hat{z} directions. We have seen this notation before,⁸ for example, in the subscripts of the Eqns. after Definition 11.1.1).

Lattices in three dimensions certainly exist where axes are non-orthogonal, but... *you will not be held responsible for any three dimensional crystal system where coordinate axes are not orthogonal.*

Two further lattice systems that you will need to know are the Face Centered Cubic (fcc) and Body Centered Cubic (bcc) lattices. In terms of our above discussion of atoms as being like small spheres, packing spheres in either a bcc or fcc lattice leaves much less open space between the spheres than packing the spheres in a simple cubic lattice.⁹ Correspondingly, these two lattices are realized much more frequently in nature.

The Body Centered Cubic (bcc) Lattice

The body centered cubic (bcc) lattice is a simple cubic lattice where there is an additional point in the very center of the cube (this is sometimes known¹⁰ as cubic-I). The unit cell is shown in the left of Fig. 11.12. Another way to show this unit cell, which does not rely on showing a three-dimensional picture, is to use a so-called *plan view* of the unit cell, shown in the right of Fig. 11.12. A plan view (a term used in engineering and architecture) is a two dimensional projection from the top of an object where heights are labeled to show the third dimension. In the picture of the bcc unit cell, there are eight lattice points on the corners of the cell (each of which is $1/8$ inside of the conventional unit cell) and one point in the center of the cell. Thus the conventional unit cell contains exactly two ($= 8 \times 1/8 + 1$) lattice points.

Packing together these unit cells to fill space, we see that the lattice points of a full bcc lattice can be described as being points having coordinates $[x, y, z]$ where either all three coordinates are integers $[uvw]$ times the lattice constant a , or all three are odd-half-integers times the lattice

⁸Note that this notation is also sometimes abused, as in Eq. 11.2, where the brackets $[a_1, a_2]$ enclose not integers, but distances which are integer multiples of a lattice constant a . To try to make things more clear, in the latter usage we will put commas between the entries, whereas the typical $[uvw]$ usage has no commas. However, most references will be extremely lax and switch between various types of notation freely.

⁹In fact it is impossible to pack spheres more densely than you would get by placing the spheres at the vertices of an fcc lattice. This result (known empirically to people who have tried to pack oranges in a crate) was first officially conjectured by Johannes Kepler in 1611, but was not mathematically proven until 1998!

¹⁰Cubic-I comes from "Innenzentriert" (inner centered). This notation was introduced by Bravais in his 1848 treatise (interestingly, Europe was burning in 1848, but obviously that didn't stop science from progressing).

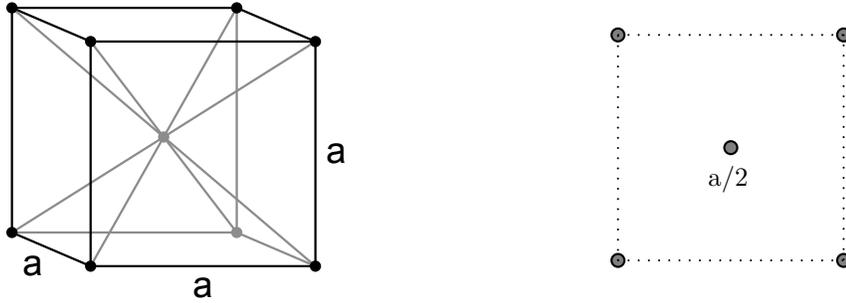


Figure 11.12: Conventional unit cell for the body centered cubic (I) lattice. Left: 3D view. Right: A plan view of the conventional unit cell. Unlabeled points are both at heights 0 and a .

constant a .

It is often convenient to think of the bcc lattice as a simple cubic lattice with a basis of two atoms per conventional cell. The simple cubic lattice contains points $[x, y, z]$ where all three coordinates are integers in units of the lattice constant. Within the conventional simple-cubic unit cell we put one point at position $[0, 0, 0]$ and another point at the position $[a/2, a/2, a/2]$ in units of the lattice constant. Thus the points of the bcc lattice are written as

$$\begin{aligned}\mathbf{R}_{\text{corner}} &= [an_1, an_2, an_3] \\ \mathbf{R}_{\text{center}} &= [an_1, an_2, an_3] + [a/2, a/2, a/2]\end{aligned}$$

as if the two different types of points were two different types of atoms, although all points in this lattice should be considered equivalent (they only look inequivalent because we have chosen a conventional unit cell with two lattice points in it).

Now, we may ask why it is that this set of points forms a lattice. In terms of our first definition of a lattice (Definition 11.1.1) we can write the primitive lattice vectors of the bcc lattice as

$$\begin{aligned}\mathbf{a}_1 &= [a, 0, 0] \\ \mathbf{a}_2 &= [0, a, 0] \\ \mathbf{a}_3 &= \left[\frac{a}{2}, \frac{a}{2}, \frac{a}{2}\right]\end{aligned}$$

It is easy to check that any combination

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3 \quad (11.4)$$

with n_1, n_2 and n_3 integers gives a point within our definition of the bcc lattice (that the three coordinates are either all integer or all half-odd integer times the lattice constant). Further one can check that any point satisfying the conditions for the bcc lattice can be written in the form of Eq. 11.4.

We can also check that our description of a bcc lattice satisfies our second description of a lattice (definition 11.1.1.1) that addition of any two points of the lattice (given by Eq. 11.4) gives another point of the lattice.

More qualitatively we can consider definition 11.1.1.2 of the lattice — that the local environment of every point in the lattice should be the same. Examining the point in the center of the unit cell, we see that it has precisely 8 nearest neighbors in each of the possible diagonal directions. Similarly any of the points in the corners of the unit cells will have 8 nearest neighbors corresponding to the points in the center of the 8 adjacent unit cells.

The *coordination number* of a lattice (frequently called Z or z) is the number of nearest neighbors any point of the lattice has. For the bcc lattice the coordination number is $Z = 8$.

As in two dimensions, a Wigner-Seitz cell can be constructed around each lattice point which encloses all points in space that are closer to that lattice point than to any other point in the lattice. This Wigner-Seitz unit cell for the bcc lattice is shown in Figure 11.13. Note that this cell is bounded by the perpendicular bisecting planes between lattice points.

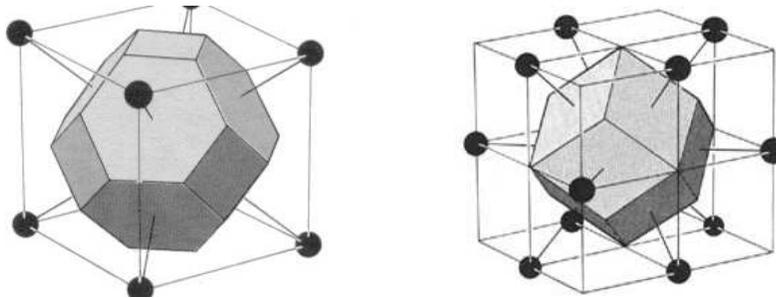


Figure 11.13: Wigner-Seitz unit cell for the bcc lattice (left) and the fcc lattice (right).

The Face Centered Cubic (fcc) Lattice

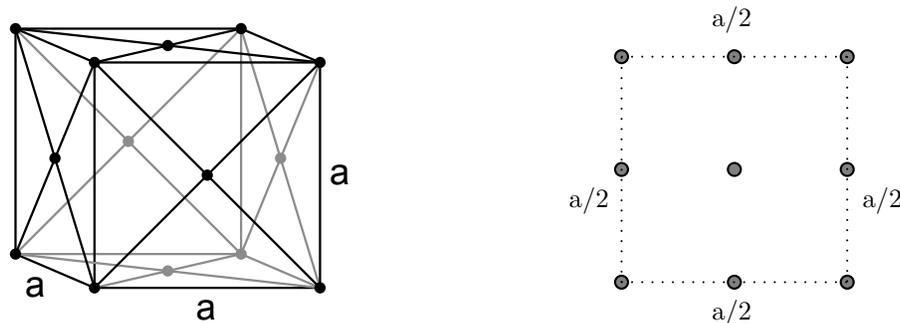


Figure 11.14: Conventional unit cell for the face centered cubic (F) lattice. Left: 3D view. Right: A plan view of the conventional unit cell. Unlabeled points are both at heights 0 and a .

The face centered (fcc) lattice is a simple cubic lattice where there is an additional point in the center of every face of every cube (this is sometimes known as cubic-F, for “face centered”). The unit cell is shown in the left of Fig. 11.14. A plan view of the unit cell is shown on the right of Fig. 11.14.

In the picture of the fcc unit cell, there are eight lattice points on the corners of the cell (each of which is $1/8$ inside of the conventional unit cell) and one point in the center of each of the 6 faces, which is $1/2$ inside the cell. Thus the conventional unit cell contains exactly four ($= 8 \times 1/8 + 6 \times 1/2$) lattice points. Packing together these unit cells to fill space, we see that the lattice points of a full fcc lattice can be described as being points having coordinates (x, y, z) where either all three coordinates are integers times the lattice constant a , or two of the three coordinates are odd-half-integers times the lattice constant a and the remaining one coordinate is an integer times the lattice constant a . Analogous to the bcc case, it is sometimes convenient to think of the fcc lattice as a simple cubic lattice with a basis of four atoms per conventional cell. The simple cubic lattice contains points $[x, y, z]$ where all three coordinates are integers in units of the lattice constant a . Within the conventional simple-cubic-unit cell we put one point at position $[0, 0, 0]$ and another point at the position $[a/2, a/2, 0]$ another at $[a/2, 0, a/2]$ and another at $[0, a/2, a/2]$. Thus the points of the fcc lattice are written as

$$\begin{aligned} \mathbf{R}_{corner} &= [an_1, an_2, an_3] \\ \mathbf{R}_{face-xy} &= [an_1, an_2, an_3] + [a/2, a/2, 0] \\ \mathbf{R}_{face-xz} &= [an_1, an_2, an_3] + [a/2, 0, a/2] \\ \mathbf{R}_{face-yz} &= [an_1, an_2, an_3] + [0, a/2, a/2] \end{aligned} \quad (11.5)$$

Again, this expresses the points of the lattice as if they were four different types of points but they only look inequivalent because we have chosen a conventional unit cell with four lattice points in it.

Again we can check that this set of points forms a lattice. In terms of our first definition of a lattice (Definition 11.1.1) we write the primitive lattice vectors of the fcc lattice as

$$\begin{aligned} \mathbf{a}_1 &= \left[\frac{a}{2}, \frac{a}{2}, 0\right] \\ \mathbf{a}_2 &= \left[\frac{a}{2}, 0, \frac{a}{2}\right] \\ \mathbf{a}_3 &= \left[0, \frac{a}{2}, \frac{a}{2}\right] \end{aligned}$$

Again it is easy to check that any combination

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$$

with n_1, n_2 and n_3 integers gives a point within our definition of the fcc lattice (that either the three coordinates are either all integer, or two of three are half-odd-integer and the remaining is integer in units of the lattice constant a).

We can also similarly check that our description of a fcc lattice satisfies our other two definitions of (definition 11.1.1.1 and 11.1.1.2) of a lattice¹¹. The Wigner-Seitz unit cell for the fcc lattice is shown in Figure 11.13.

Other Lattices in Three Dimensions

In addition to the simple cubic, orthorhombic, tetragonal, fcc, and bcc lattices, there are nine other types of lattices in three dimensions. These are known as the fourteen *Bravais lattice types*¹² You

¹¹Can you figure out the coordination number of the fcc lattice? Find the minimum distance between two lattice points then find out how many lattice points are this distance away from any one given lattice point. (It would be too easy if I told you the answer!)

¹²Named after Auguste Bravais who classified all the three dimensional lattices in 1848. Actually they should be named after Moritz Frankenheim who studied the same thing over ten years earlier — although he made a minor

are not responsible for knowing these! But it is probably a good idea to know that they exist.

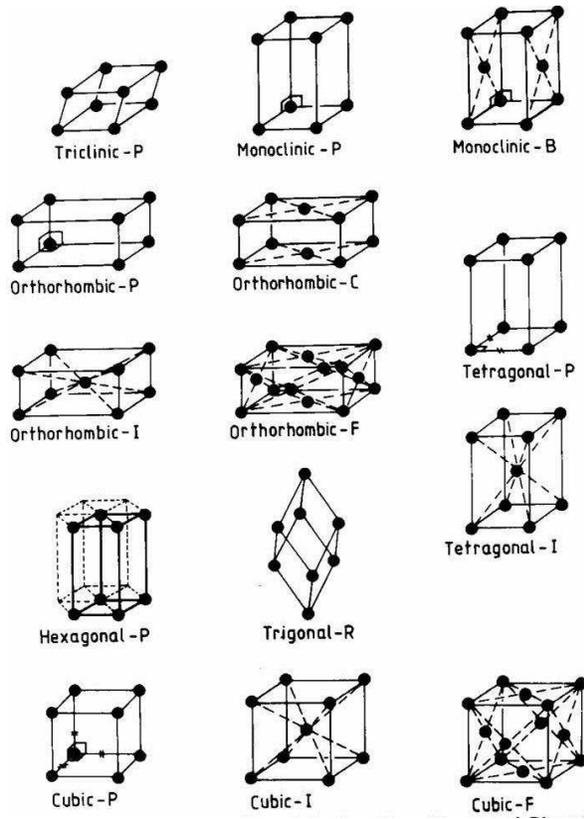


Figure 11.15: Unit cells for All of the Three Dimensional Bravais Lattice Types.

Figure 11.15 shows the full variety of Bravais lattice types in three dimensions. While it is an extremely deep fact that there are only 14 lattice types in three dimensions, the precise statement of this theorem, as well of the proof of it, are beyond the scope of this course. The key result is that any crystal, no matter how complicated, has a lattice which is one of these 14 types.¹³

Real Crystals

Once we have discussed lattices we can combine a lattice with a basis to describe any periodic structure — and in particular, we can describe any crystalline structure.

error in his studies, and therefore missed getting his name associated with them.

¹³There is a real subtlety here in classifying a crystal as having a particular lattice type. There are only these 14 lattice types, but in principle a crystal could have one lattice, but have the *symmetry* of another lattice. An example of this would be if the a lattice were cubic, but the unit cell did not look the same from all six sides. Crystallographers would not classify this as being a cubic material even if the lattice happened to be cubic. The reason for this is that if the unit cell did not look the same from all six sides, there would be no particular reason that the three primitive lattice vectors should have the same length — it would be an insane coincidence were this to happen, and almost certainly in any real material the primitive lattice vector lengths would actually have slightly different values if measured more closely.

Several examples of real (and reasonably simple) crystal structures are shown in Fig. 11.16.

11.3 Summary of Crystal Structure

This chapter introduced a plethora of new definitions, aimed at describing crystal structure in three dimensions. Here is a list of some of the concepts that one should know

- Definition of a *lattice* (in three different ways See definitions 11.1.1, 11.1.1.1, 11.1.1.2)
- Definition of a *unit cell* for a periodic structure, and definition of a *primitive unit cell* and a *conventional unit cell*
- Definition and construction of the *Wigner-Seitz (primitive) unit cell*.
- One can write any periodic structure in terms of a lattice and a basis (See examples in Fig. 11.16).
- In 3d, know the simple cubic lattice, the fcc lattice and the bcc lattice.
- The fcc and bcc lattices can be thought of as simple cubic lattices with a basis.
- Know how to read a *plan view* of a structure.

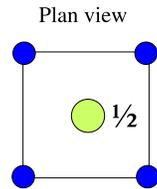
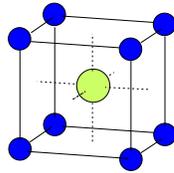
References

All books cover this. Some books give *way* too much detail for us. I recommend the following as giving not too much and not too little:

- Kittel, chapter 1
- Ashcroft and Mermin chapter 4 (caution of the nomenclature issue, see footnote 1 of this chapter).

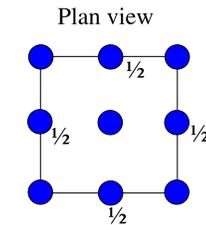
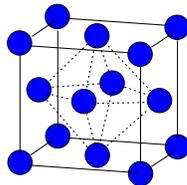
cesium chloride (CsCl)

lattice: cubic P
 basis :
 Cs 000 ●
 Cl $\frac{1}{2}\frac{1}{2}\frac{1}{2}$ ●



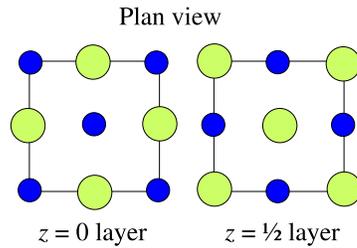
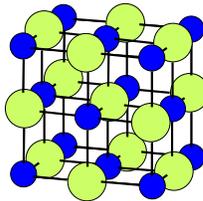
copper (Cu)

lattice: cubic F
 basis :
 Cu 000 ●



sodium chloride (NaCl)

lattice: cubic F
 basis :
 Na 000 ●
 Cl $\frac{1}{2}\frac{1}{2}\frac{1}{2}$ ●



diamond (C) — also Si, Ge

lattice: cubic F
 basis :
 C 000 ●
 C $\frac{1}{4}\frac{1}{4}\frac{1}{4}$ ●

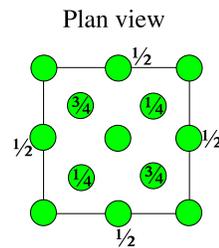
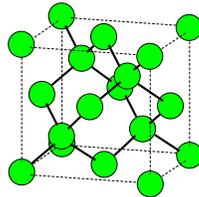


Figure 11.16: Some examples of real crystals with simple structures. Note that in all cases the basis is described with respect to the primitive unit cell of a simple cubic lattice.

Chapter 12

Reciprocal Lattice, Brillouin Zone, Waves in Crystals

In the last chapter we explored lattices and crystal structure. However as we saw in chapters 8–10, the important physics of waves in solids (whether they be vibrational waves, or electron waves) is best described in reciprocal space. This chapter thus introduces reciprocal space in 3 dimensions. As with the previous chapter, there is some tricky geometry in this chapter, and a few definitions to learn as well. This makes this material a bit tough to slog through, but stick with it because soon we will make substantial use of what we learn here. At the end of this chapter we will finally have enough definitions to describe the dispersions of phonons and electrons in three dimensional systems.

12.1 The Reciprocal Lattice in Three Dimensions

12.1.1 Review of One Dimension

Let us first recall some results from our study of one dimension. We consider a simple lattice in one dimension $R_n = na$ with n an integer. Recall that two points in k -space (reciprocal space) were defined to be equivalent to each other if $k_1 = k_2 + G_m$ where $G_m = 2\pi m/a$ with m an integer. The points G_m form the reciprocal lattice.

Recall that the reason that we identified different k values was because we were considering waves of the form

$$e^{ikx_n} = e^{ikna}$$

with n an integer. Because of this form of the wave, we find that shifting $k \rightarrow k + G_m$ leaves this functional form unchanged since

$$e^{i(k+G_m)x_n} = e^{i(k+G_m)na} = e^{ikna} e^{i(2\pi m/a)na} = e^{ikx_n}$$

where we have used

$$e^{i2\pi mn} = 1$$

in the last step. Thus, so far as the wave is concerned, k is the same as $k + G_m$.

12.1.2 Reciprocal Lattice Definition

Generalizing the above result from one dimension, we make the following definition:

Definition 12.1.1. Given a (direct) lattice of points \mathbf{R} , a point \mathbf{G} is a point in the reciprocal lattice if and only if

$$e^{i\mathbf{G}\cdot\mathbf{R}} = 1 \quad (12.1)$$

for all points \mathbf{R} of the direct lattice.

To construct the reciprocal lattice, let us write the points of the direct lattice in the form (Here we specialize to the three dimensional case).

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3 \quad (12.2)$$

with n_1, n_2 and n_3 integers, and with $\mathbf{a}_1, \mathbf{a}_2$, and \mathbf{a}_3 being primitive lattice vectors of the direct lattice.

We now make two key claims:

1. We claim that the reciprocal lattice (defined by Eq. 12.1) is a lattice in reciprocal space (thus explaining its name).
2. We claim that the primitive lattice vectors of the reciprocal lattice (which we will call $\mathbf{b}_1, \mathbf{b}_2$, and \mathbf{b}_3) are defined to have the following property:

$$\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij} \quad (12.3)$$

where δ_{ij} is the Kronecker delta¹.

We can certainly construct vectors \mathbf{b}_i to have the desired property of Eq. 12.3, as follows:

$$\begin{aligned} \mathbf{b}_1 &= \frac{2\pi \mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)} \\ \mathbf{b}_2 &= \frac{2\pi \mathbf{a}_3 \times \mathbf{a}_1}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)} \\ \mathbf{b}_3 &= \frac{2\pi \mathbf{a}_1 \times \mathbf{a}_2}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)} \end{aligned}$$

It is easy to check that Eq. 12.3 is satisfied. For example,

$$\begin{aligned} \mathbf{a}_1 \cdot \mathbf{b}_1 &= \frac{2\pi \mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)} = 2\pi \\ \mathbf{a}_2 \cdot \mathbf{b}_1 &= \frac{2\pi \mathbf{a}_2 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)} = 0 \end{aligned}$$

Now, given vectors $\mathbf{b}_1, \mathbf{b}_2$, and \mathbf{b}_3 satisfying Eq. 12.3 we have claimed that these are in fact primitive lattice vectors for the reciprocal lattice.

¹Leopold Kronecker was a mathematician who is famous (among other things) for the sentence “God made the integers, everything else is the work of man”. In case you don’t already know this, the Kronecker delta is defined as $\delta_{ij} = 1$ for $i = j$ and is zero otherwise. (Kronecker did a lot of other interesting things as well)

Let us write an arbitrary point in reciprocal space as

$$\mathbf{G} = m_1 \mathbf{b}_1 + m_2 \mathbf{b}_2 + m_3 \mathbf{b}_3 \quad (12.4)$$

and for the moment, let us not require m_1, m_2 and m_3 to be integers. (We are about to discover that for \mathbf{G} to be a point of the reciprocal lattice, they must be integers, but this is what we want to prove!).

To find points of the reciprocal lattice we must show that Eq. 12.1 is satisfied for all points $\mathbf{R} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3$ of the direct lattice with n_1, n_2 and n_3 integers. We thus write

$$e^{i\mathbf{G}\cdot\mathbf{R}} = e^{i(m_1 \mathbf{b}_1 + m_2 \mathbf{b}_2 + m_3 \mathbf{b}_3) \cdot (n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3)} = e^{2\pi i(n_1 m_1 + n_2 m_2 + n_3 m_3)}$$

In order for \mathbf{G} to be a point of the reciprocal lattice, this must equal unity for all points \mathbf{R} of this direct lattice, i.e., for all integer values of n_1, n_2 and n_3 . Clearly this can only be true if m_1, m_2 and m_3 are also integers. Thus, we find that the points of the reciprocal lattice are precisely those of the form of Eq. 12.4 with m_1, m_2 and m_3 integers. This further proves our claim that the reciprocal lattice is in fact a lattice!

12.1.3 The Reciprocal Lattice as a Fourier Transform

Quite generally one can think of the Reciprocal lattice as being a Fourier transform of the direct lattice. It is easiest to start by thinking in one dimension. Here the direct lattice is given again by $R_n = an$. If we think of the “density” of lattice points in one dimension, we might put a delta-function of density at these lattice points so we write the density as²

$$\rho(r) = \sum_n \delta(r - an)$$

Fourier transforming this function gives³

$$\mathcal{F}[\rho(r)] = \int dr e^{ikr} \rho(r) = \sum_n \int dr e^{ikr} \delta(r - an) = \sum_n e^{ikan} = 2\pi \sum_m \delta(k - 2\pi m/a)$$

The last step here is a bit nontrivial.⁴ Here e^{ikan} is clearly unity if $k = 2\pi m/a$, i.e., if k is a point on the reciprocal lattice. In this case, each term of the sum contributes unity to the sum and one obtains an infinite result. If k is not such a reciprocal lattice point, then the terms of the sum oscillate and the sum comes out to be zero.

This principle generalizes to the higher (two and three) dimensional cases. Generally

$$\mathcal{F}[\rho(\mathbf{r})] = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} = (2\pi)^D \sum_{\mathbf{G}} \delta^D(\mathbf{k} - \mathbf{G}) \quad (12.5)$$

²Since the sums are over all lattice points they should go from $-\infty$ to $+\infty$. Alternately, one uses periodic boundary conditions and sums over all points.

³With Fourier transforms there are many different conventions about where one puts the factors of 2π . Probably in your mathematics class you learned to put $1/\sqrt{2\pi}$ with each k integral and with each r integral. However, in Solid-State physics conventionally $1/(2\pi)$ comes with each k integral, and no factor of 2π comes with each r integral. See section 2.2.1 to see why this is used.

⁴This is sometimes known as the Poisson resummation formula, after Siméon Denis Poisson, the same guy after whom Poisson’s equation $\nabla^2 \phi = -\rho/\epsilon_0$ is named, as well as other mathematical things such as the Poisson random distribution. His last name means “fish” in French.

where in the middle term, the sum is over lattice points \mathbf{R} of the direct lattice, and in the last term it is a sum over points \mathbf{G} of the reciprocal lattice. Here D is the number of dimensions (1, 2 or 3) and the δ^D is a D -dimensional delta function⁵. This equality is similar to that explained above. As above, if \mathbf{k} is a point of the reciprocal lattice, then $e^{i\mathbf{k}\cdot\mathbf{R}}$ is always unity and the sum is infinite (a delta function). However, if \mathbf{k} is not a point on the reciprocal lattice then the summands oscillate, and the sum comes out to be zero. Thus one obtains delta function peaks precisely at the positions of reciprocal lattice vectors.

Aside: It is an easy exercise to show⁶ that the reciprocal lattice of an fcc direct lattice is a bcc lattice in reciprocal space. Conversely, the reciprocal lattice of a bcc direct lattice is an fcc lattice in reciprocal space.

Fourier Transform of Any Periodic Function

In the above section we considered the Fourier transform of a function $\rho(\mathbf{r})$ which is just a set of delta functions at lattice points. However, it is not too different to consider the Fourier transform of any function with the periodicity of the lattice (and this will be quite important below in chapter 13). We say a function $\rho(\mathbf{r})$ has the periodicity of a lattice if periodic $\rho(\mathbf{r}) = \rho(\mathbf{r} + \mathbf{R})$ for any lattice vector \mathbf{R} . We then want to calculate

$$\mathcal{F}[\rho(\mathbf{r})] = \int d\mathbf{r} e^{i\mathbf{k}\cdot\mathbf{r}} \rho(\mathbf{r})$$

The integral over all of space can be broken up into a sum of integrals over each unit cell. Here we write any point in space \mathbf{r} as the sum of a lattice point \mathbf{R} and a vector \mathbf{x} within the unit cell

$$\mathcal{F}[\rho(\mathbf{r})] = \sum_{\mathbf{R}} \int_{unit-cell} d\mathbf{x} e^{i\mathbf{k}\cdot(\mathbf{x}+\mathbf{R})} \rho(\mathbf{x} + \mathbf{R}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \int_{unit-cell} d\mathbf{x} e^{i\mathbf{k}\cdot\mathbf{x}} \rho(\mathbf{x})$$

where here we have used the invariance of ρ under lattice translations $\mathbf{x} \rightarrow \mathbf{x} + \mathbf{R}$. The first term, as in Eq. 12.5 just gives a sum of delta functions yielding

$$\mathcal{F}[\rho(\mathbf{r})] = (2\pi)^D \sum_{\mathbf{G}} \delta^D(\mathbf{k} - \mathbf{G}) S(\mathbf{k})$$

where

$$S(\mathbf{k}) = \int_{unit-cell} d\mathbf{x} e^{i\mathbf{k}\cdot\mathbf{x}} \rho(\mathbf{x}) \quad (12.6)$$

is known as the *structure factor* and will become very important in the next chapter.

12.1.4 Reciprocal Lattice Points as Families of Lattice Planes

Another way to understand the reciprocal lattice is via families of lattice planes of the direct lattice.

Definition 12.1.2. A *lattice plane* (or *crystal plane*) is a plane containing at least three noncolinear (and therefore an infinite number of) points of a lattice.

Definition 12.1.3. A *family of lattice planes* is an infinite set of equally separated lattice planes which taken together contain all points of the lattice.

⁵For example, in two dimensions $\delta^2(\mathbf{r} - \mathbf{r}_0) = \delta(x - x_0)\delta(y - y_0)$ where $\mathbf{r} = (x, y)$

⁶Try it!

In Figure 12.1, two examples of families of lattice planes are shown. Note that the planes are parallel and equally spaced, and every point of the lattice is included in exactly one lattice plane.

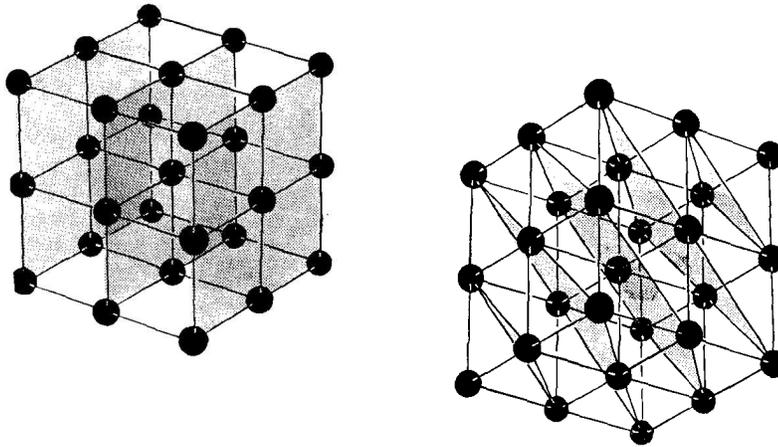


Figure 12.1: Two Examples of Families of Lattice planes on the Cubic Lattice. Each of these planes is a crystal plane because it intersects an infinite number of lattice points. The left example is (100) and the right example is (111) in the Miller index notation.

I now make the following claim:

Claim: *The families of lattice planes are in one-to-one correspondence⁷ with the possible directions of reciprocal lattice vectors, to which they are normal. Further the spacing between these lattice planes is $d = 2\pi/|\mathbf{G}_{\min}|$ where \mathbf{G}_{\min} is the minimum length reciprocal lattice vector in this normal direction.*

This correspondence is made as follows. First we consider the set of planes defined by points \mathbf{r} such that

$$\mathbf{G} \cdot \mathbf{r} = 2\pi m \quad (12.7)$$

This defines an infinite set of parallel planes normal to \mathbf{G} . Since $e^{i\mathbf{G} \cdot \mathbf{r}} = 1$ we know that every lattice point is a member of one of these planes (since this is the definition of \mathbf{G} in Eq. 12.1). However, for the planes defined by Eq. 12.7, not every plane needs to contain a lattice point (so generically this is a family of parallel equally spaced planes, but not a family of lattice planes). For this larger family of planes, the spacing between planes is given by

$$d = \frac{2\pi}{|\mathbf{G}|} \quad (12.8)$$

⁷For this one-to-one correspondence to be precisely true we must define \mathbf{G} and $-\mathbf{G}$ to be the same direction. If this sounds like a cheap excuse, we can say that “oriented” families of lattice planes are in one-to-one correspondence with the directions of reciprocal lattice vectors, thus keeping track of the two possible normals of the family of lattice planes.

To prove this we simply note that two adjacent planes must have

$$\mathbf{G} \cdot (\mathbf{r}_1 - \mathbf{r}_2) = 2\pi$$

Thus in the direction parallel to \mathbf{G} , the spacing between planes is $2\pi/|\mathbf{G}|$ as claimed.

Clearly different values of \mathbf{G} that happen to point in the same direction, but have different magnitudes, will define parallel sets of planes. As we increase the magnitude of \mathbf{G} , we add more and more lattice planes. For example, examining Eq. 12.7 we see that when we double the magnitude of \mathbf{G} we correspondingly double the density of planes, which we can see from the spacing formula Eq. 12.8. However, whichever \mathbf{G} we choose, all of the lattice points will be included in one of the defined planes. If we choose the maximally possible spaced planes, hence the smallest possible value of \mathbf{G} allowed in any given direction which we call \mathbf{G}_{\min} , then in fact every defined plane will include lattice points and therefore be lattice planes, and the spacing between these planes is correspondingly $2\pi/|\mathbf{G}_{\min}|$.⁸ This proves our above claim.

12.1.5 Lattice Planes and Miller Indices

It is convenient to define a notation for describing lattice planes. The conventional notations are known as *Miller Indices*.⁹ One writes

$$(h, k, l) \quad \text{or} \quad (hkl)$$

with integers h, k and l , to mean a family of lattice planes corresponding to reciprocal lattice vector

$$\mathbf{G}_{(h,k,l)} = h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3 \quad (12.9)$$

where \mathbf{b}_i are the standard primitive lattice vectors of the reciprocal lattice¹⁰. Note that (h, k, l) as a family of lattice planes, should be the shortest reciprocal lattice vector in that direction, meaning that the integers h, k and l should have no common divisor. One may also write (h, k, l) where h, k and l do have a common divisor, but then one is talking about a reciprocal lattice vector, or a family of planes that is not a family of lattice planes (i.e., there are some planes that do not intersect lattice point).

Important Complication: For fcc and bcc lattices, Miller indices are usually stated using the primitive lattice vectors of the *cubic* lattice in Eq. 12.9 rather than the primitive lattice vector of the fcc or bcc.

This comment is quite important. For example, the (100) family of planes for the cubic lattice (shown in the right of Fig. 12.1) intersects every corner of the cubic unit cell. However, if we were discussing a bcc lattice, there would also be another lattice point in the center of every conventional unit cell, and the (100) lattice planes would not intersect. However, the (200) planes would intersect these central points as well, so in this case (200) represents a true family of lattice planes for the bcc lattice whereas (100) does not!

⁸More rigorously, if there is a family of lattice planes in direction $\hat{\mathbf{G}}$ with spacing between planes d , then $\mathbf{G} = 2\pi\hat{\mathbf{G}}/d$ is necessarily a reciprocal lattice vector. To see this note that $e^{i\mathbf{G}\cdot\mathbf{R}} = 1$ will be unity for all lattice points. Further, in a family of lattice planes, all lattice points are included within the planes, so $e^{i\mathbf{G}\cdot\mathbf{R}} = 1$ for all \mathbf{R} a lattice point, which implies \mathbf{G} is a reciprocal lattice vector. Furthermore, \mathbf{G} is the shortest reciprocal lattice vector in the direction of $\hat{\mathbf{G}}$ since increasing \mathbf{G} will result in a smaller spacing of lattice planes and some planes will not intersect lattice points \mathbf{R} .

⁹These are named after the 19th century mineralogist William Hallows Miller. It is interesting that the structure of lattice planes was understood long before the world was even certain there was such a thing as an atom.

¹⁰We have already used the corresponding notation $[uvw]$ to represent lattice points of the direct lattice. See for example, Eq. 11.1 and Eq. 11.3.

Figure 9 This plane intercepts the a , b , c axes at $3a$, $2b$, $2c$. The reciprocals of these numbers are $\frac{1}{3}, \frac{1}{2}, \frac{1}{2}$. The smallest three integers having the same ratio are 2, 3, 3, and thus the Miller indices of the plane are (233).

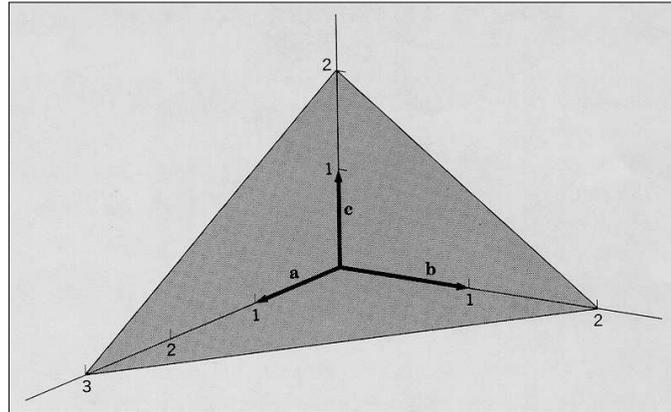


Figure 12.2: Determining Miller Indices From the Intersection of a Plane with the Coordinate Axes. The spacing between lattice planes in this family would be $\frac{1}{|d_{(233)}|^2} = \frac{2^2}{a^2} + \frac{3^2}{b^2} + \frac{3^2}{c^2}$.

From Eq. 12.8 one can write the spacing between a family of planes specified by Miller indices (h, k, l)

$$d_{(hkl)} = \frac{2\pi}{|\mathbf{G}|} = \frac{2\pi}{\sqrt{h^2|\mathbf{b}_1|^2 + k^2|\mathbf{b}_2|^2 + l^2|\mathbf{b}_3|^2}} \quad (12.10)$$

where we have assumed that the coordinate axes of the primitive lattice vectors \mathbf{b}_i are orthogonal. Recall that in the case of orthogonal axes $|\mathbf{b}_i| = 2\pi/a_i$ where a_i are the lattice constants in the three orthogonal directions. Thus we can equivalently write

$$\frac{1}{|d_{(hkl)}|^2} = \frac{h^2}{a_1^2} + \frac{k^2}{a_2^2} + \frac{l^2}{a_3^2} \quad (12.11)$$

Note that for a cubic lattice this simplifies to

$$d_{(hkl)}^{cubic} = \frac{a}{\sqrt{h^2 + k^2 + l^2}} \quad (12.12)$$

A useful shortcut for figuring out the geometry of lattice planes is to look at the intersection of a plane with the three coordinate axes. The intersections x_1, x_2, x_3 with the three coordinate

axes (in units of the three principle lattice constants) are related to the Miller indices via

$$\frac{a_1}{x_1} : \frac{a_2}{x_2} : \frac{a_3}{x_3} = h : k : l$$

This construction is illustrated in Fig. 12.2.

In figure Fig. 12.3 we show three more examples of Miller indices.

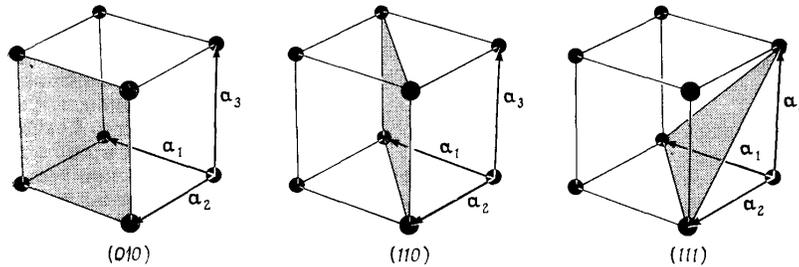


Figure 12.3: More Examples of Miller Indices.

Note that Miller indices can be negative if the planes intersect the negative axes. We could have, for example, a lattice plane (1,-1,1). Conventionally, the minus sign is denoted with an over-bar rather than a minus sign, so we write $(1\bar{1}1)$ instead¹¹.

Finally, we note that different lattice planes may be the same under a symmetry of the crystal. For example, in a cubic lattice, (111) looks the same as $(1\bar{1}\bar{1})$ after rotation (and possibly reflection) of the axes of the crystal (but would never look like (122) under any rotation or reflection since the spacing between planes is different!). If we want to describe all lattice planes that are equivalent in this way, we write $\{111\}$ instead.

It is interesting that lattice planes in crystals were well understood long before people even knew for sure there was such a thing as atoms. By studying how crystals cleave along certain

¹¹How $(1\bar{1}\bar{1})$ is pronounced is a bit random. Some people say “one-(bar-one)-one” and others say “one-(one-bar)-one”. I have no idea how the community got so confused as to have these two different conventions. I think in Europe the former is more prevalent whereas in America the latter is more prevalent. At any rate, it is always clear when it is written.

planes, scientists like Miller and Bravais could reconstruct a great deal about how these materials must be assembled¹².

12.2 Brillouin Zones

The whole point of going into such gross detail about the structure of reciprocal space is in order to describe waves in solids. In particular, it will be important to understand the structure of the Brillouin zone.

12.2.1 Review of One Dimensional Dispersions and Brillouin Zones

As we learned in chapters 8–10, the Brillouin zone is extremely important in describing the excitation spectrum of waves in periodic media. As a reminder, in Fig. 12.4 we show the excitation spectrum of vibrations of a diatomic chain (chapter 9) in both the reduced, and extended zone schemes. Since waves are physically equivalent under shifts of the wavevector k by a reciprocal lattice vector $2\pi/a$, we can always express every excitation within the first Brillouin zone, as shown in the reduced zone scheme (left of Fig. 12.4). In this example, since there are two atoms per unit cell, there are precisely two excitation modes per wavevector. On the other hand, we can always unfold the spectrum and put the lowest (acoustic) excitation mode in the first Brillouin zone and the higher energy excitation mode (optical) in the second Brillouin zone, as shown in the extended zone scheme (right of Fig. 12.4). Note that there is a jump in the excitation spectrum at the Brillouin zone boundary.

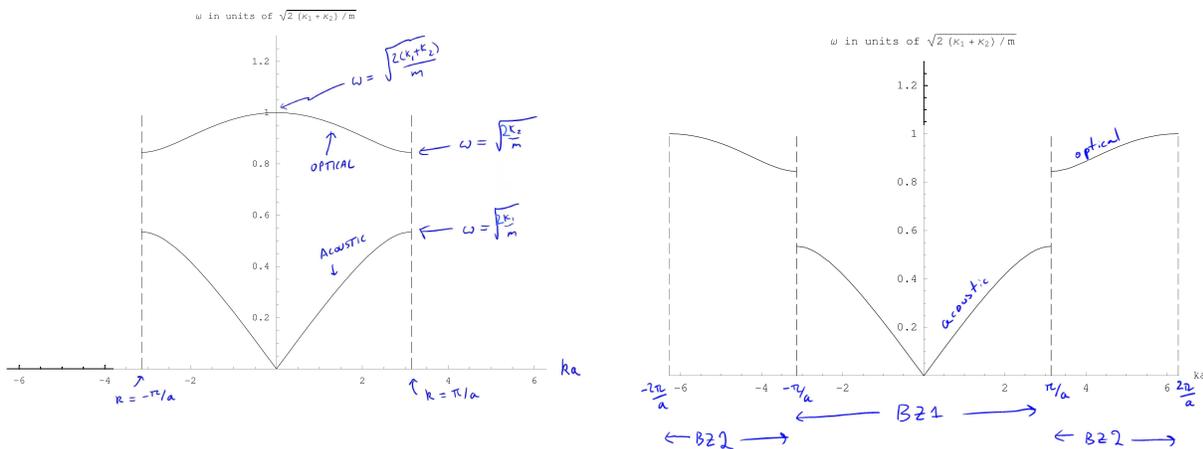


Figure 12.4: Phonon Spectrum of a Diatomic Chain in One Dimension. Left: Reduced Zone scheme. Right: Extended Zone scheme. (See Figs. 9.1 and 9.2)

¹²There is a law known as “Bravais’ Law” which states that crystals cleave most readily along faces having the highest density of lattice points. In modern language this is essentially equivalent to stating that the fewest atomic bonds should be broken in the cleave. Can you see why this is?

12.2.2 General Brillouin Zone Construction

Definition 12.2.1. A *Brillouin zone* is a unit cell of the reciprocal lattice.

Entirely equivalent to the one dimensional situation, physical waves in crystals are unchanged if their wavevector is shifted by a reciprocal lattice vector $\mathbf{k} \rightarrow \mathbf{k} + \mathbf{G}$. Alternately, we realize that the physically relevant quantity is the crystal momentum. Thus, the Brillouin zone has been defined to include each physically different crystal momentum exactly once (Each \mathbf{k} point within the Brillouin zone is physically different and all physically different points occur once within the zone).

While the most general definition of Brillouin zone allows us to choose any shape unit cell for the reciprocal lattice, there are some definitions of unit cells which are more convenient than others.

We define the *first Brillouin zone* in reciprocal space quite analogously to the construction of the Wigner-Seitz cell for the direct lattice.

Definition 12.2.2. Start with the reciprocal lattice point $\mathbf{G} = \mathbf{0}$. All \mathbf{k} points which are closer to $\mathbf{0}$ than to any other reciprocal lattice point define the *first Brillouin zone*. Similarly all points where the point $\mathbf{0}$ is the *second* closest reciprocal lattice point to that point constitute the *second Brillouin zone*, and so forth. *Zone boundaries* are defined in terms of this definition of Brillouin zones.

As with the Wigner-Seitz cell, there is a simple algorithm to construct the Brillouin zones. Draw the perpendicular bisector between the point $\mathbf{0}$ and each of the reciprocal lattice vectors. These bisectors form the Brillouin zone boundaries. Any point that you can get to from $\mathbf{0}$ without crossing a reciprocal lattice vector is in the first Brillouin zone. If you cross only one perpendicular bisector, you are in the second Brillouin zone, and so forth.

In figure 12.5, we show the Brillouin zones of the square lattice. A few general principles to note:

1. The first Brillouin zone is necessarily connected, but the higher Brillouin zones typically are made of disconnected pieces.
2. A point on a Brillouin zone boundary lies on the perpendicular bisector between the point $\mathbf{0}$ and some reciprocal lattice point \mathbf{G} . Adding the vector $-\mathbf{G}$ to this point necessarily results in a point (the same distance from $\mathbf{0}$) which is on another Brillouin zone boundary (on the bisector of the segment from $\mathbf{0}$ to $-\mathbf{G}$). This means that Brillouin zone boundaries occur in parallel pairs symmetric around the point $\mathbf{0}$ which are separated by a reciprocal lattice vector (See Fig. 12.5).
3. Each Brillouin zone has exactly the same total area (or volume in three dimensions). This must be the case since there is a one-to-one mapping of points in each Brillouin zone to the first Brillouin zone. Finally, as in 1d, we claim that there are exactly as many k -states within the first Brillouin zone as there are unit cells in the entire system¹³.

Note, that as in the case of the Wigner Seitz cell construction, the shape of the first Brillouin zone can look a bit strange, even for a relatively simple lattice (See Fig. 11.7).

¹³Here's the proof of this statement for a square lattice. Let the system be N_x by N_y unit cells. Then, with

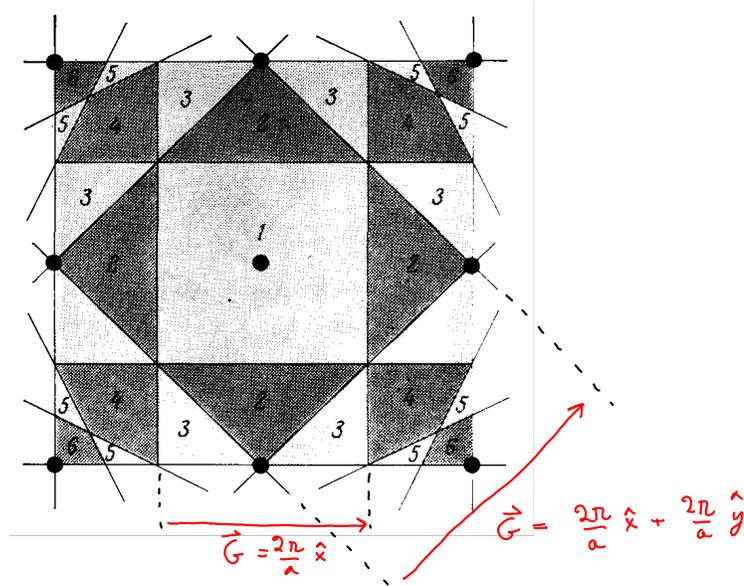


Figure 12.5: First, second, third, fourth, . . . Brillouin zones of the square lattice. Note that zone boundaries occur in parallel pairs symmetric around $\mathbf{0}$ and separated by a reciprocal lattice vector.

The construction of the Brillouin zone is similar in three dimensions as it is in two, and is again entirely analogous to the construction of the Wigner-Seitz cell in three dimensions. For a simple cubic lattice, the first Brillouin zone is simply a cube. For fcc and bcc lattices, however, the situation is more complicated. As we mentioned above in the *Aside* at the end of section 12.1.3 above, the reciprocal lattice of the fcc lattice is bcc, and vice-versa. Thus, the Brillouin zone of the fcc lattice is the same shape as the the Wigner-Seitz cell of the bcc lattice! The Brillouin zone for the fcc lattice is shown in Fig. 12.6 (compare to Fig. 11.13). Note that in Fig. 12.6, various k -points are labeled with letters. There is a complicated labeling convention that we will not discuss in this course, but it is worth knowing that it exists. For example, we can see in the figure that the point $\mathbf{k} = \mathbf{0}$ is labeled Γ , and the point $\mathbf{k} = (\pi/a)\hat{y}$ is labeled X .

Given this diagram of this Brillouin zone we can finally arrive at some real physics!

12.3 Electronic and Vibrational Waves in Crystals in Three Dimensions

In the left of Fig. 12.7 we show the electronic band-structure (i.e., dispersion relation) of diamond, which is an fcc lattice with a diatomic basis (See Fig. 11.16). As in the one-dimensional case, we can choose to work in the reduced zone scheme where we only need to consider the first Brillouin zone. Since we are trying to display a three dimensional spectrum (Energy as a function of \mathbf{k}) on a one dimensional diagram, what is done is to show several single-line cuts through reciprocal

periodic boundary conditions, the value of k_x is quantized in units of $2\pi/L_x = 2\pi/(N_x a)$ and the value of k_y is quantized in units of $2\pi/L_y = 2\pi/(N_y a)$. But the size of the Brillouin zone is $2\pi/a$ in each direction, thus there are precisely $N_x N_y$ different values of \mathbf{k} in the Brillouin zone.

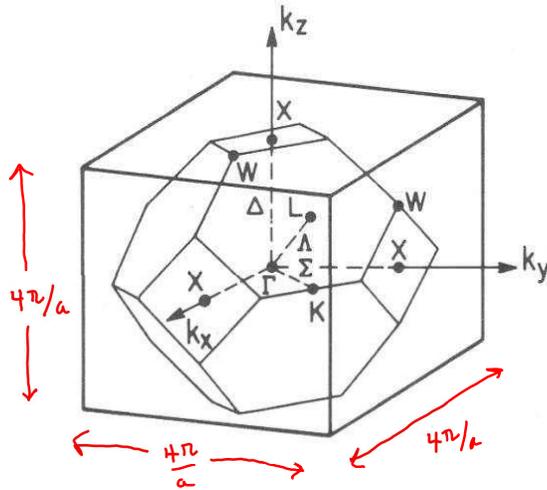


Figure 12.6: First Brillouin Zone of the FCC Lattice. Note that it is the same shape as the Wigner-Seitz cell of the bcc lattice, see Fig. 11.13. Various special points of the Brillouin zone are labeled with code letters such as X , K , and Γ .

space¹⁴. Starting on the left of the diagram, we start at L -point of the Brillouin zone and show $E(\mathbf{k})$ as \mathbf{k} traces a straight line to the Γ point (the center of the Brillouin zone). Then we continue to the right and \mathbf{k} traces a straight line from the Γ point to the X point. Note that the lowest band is quadratic at the center of the Brillouin zone (a dispersion $\hbar^2 k^2 / (2m^*)$ for some effective mass m^*).

Similarly, in the right of Fig. 12.7, we show the phonon spectrum of diamond. Several things to note about this figure. First of all, since diamond has a unit cell with two atoms in it (it is fcc with a basis of two atoms) there should be six modes of oscillation per k -points (three directions of motion times two atoms per unit cell). Indeed, this is what we see in the picture, at least in the central third of the picture. In the other two parts of the picture, one sees fewer modes per k -point, but this is because, due to the symmetry of the crystal along this particular direction, several excitation modes have exactly the same energy (you can see for example, at the X -point, two modes come in from the right, but only one goes out to the left. This means the two modes have the same energy on the left of the X point). Secondly, we note that at the Γ -point, $\mathbf{k} = 0$ there are exactly three modes which come down linearly to zero energy. These are the three acoustic modes. The other three modes, which are finite energy at $\mathbf{k} = 0$ are the optical modes. Finally, you may note something a bit confusing about this diagram. On the far left of the diagram, we start at the Γ point, move in the (100) direction and end up at the X point. Then from the X point, we move in the (110) direction, and we end up back at the Γ point! This is because we have landed at the Γ point in a different Brillouin zone.

¹⁴This type of plot, because it can look like a jumble of lines, is sometimes called a “spaghetti-diagram”

- Ashcroft and Mermin, chapter 5 (again be warned of the nomenclature issue mentioned above in chapter 11, footnote 1).

Many books introduce X-ray diffraction and the reciprocal lattice at the same time. Once we have read the next chapter and we study scattering, we might go back and look at the nice introductions to reciprocal space given in the following books

- Goodstein, section 3.4–3.5 (very brief)
- Kittel, chapter 2
- Ibach and Luth, chapter 3

Part V

Neutron and X-Ray Diffraction

Chapter 13

Wave Scattering by Crystals

In the last chapter we discussed reciprocal space, and explained that the energy dispersion of phonons and electrons is plotted within the Brillouin zone. We understand how these are similar to each other due to the wave-like nature of both the electron and the phonon. However, much of the same physics occurs when a crystal scatters waves (or particles¹) that impinge upon it externally. Indeed, exposing a solid to a wave in order to probe its properties is an extremely useful thing to do. The most commonly used probe is X-rays. Another common, more modern, probe is neutrons. It can hardly be overstated how important this type of experiment is to science.

The general setup that we will examine is shown in Fig.13.1.

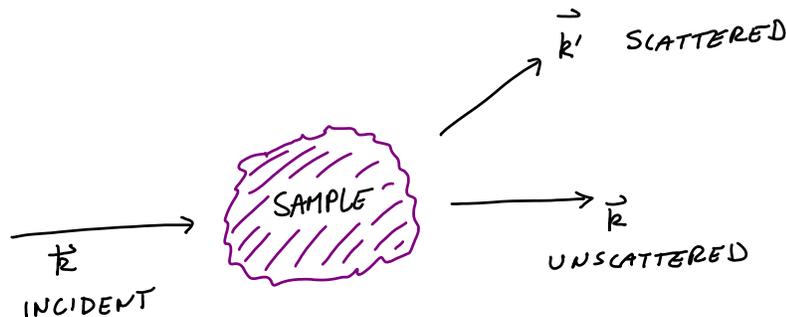


Figure 13.1: A generic scattering experiment.

¹Remember, in quantum mechanics there is no real difference between particles and waves!

13.1 The Laue and Bragg Conditions

13.1.1 Fermi's Golden Rule Approach

If we think of the incoming wave as being a particle, then we should think of the sample as being some potential $V(\mathbf{r})$ that the particle experiences as it goes through the sample. According to Fermi's golden rule², the transition rate $\Gamma(\mathbf{k}', \mathbf{k})$ per unit time for the particle scattering from \mathbf{k} to \mathbf{k}' is given by

$$\Gamma(\mathbf{k}', \mathbf{k}) = \frac{2\pi}{\hbar} |\langle \mathbf{k}' | V | \mathbf{k} \rangle|^2 \delta(E_{\mathbf{k}'} - E_{\mathbf{k}})$$

The matrix element here

$$\langle \mathbf{k}' | V | \mathbf{k} \rangle = \int d\mathbf{r} \frac{e^{-i\mathbf{k}' \cdot \mathbf{r}}}{\sqrt{L^3}} V(\mathbf{r}) \frac{e^{i\mathbf{k} \cdot \mathbf{r}}}{\sqrt{L^3}} = \frac{1}{L^3} \int d\mathbf{r} e^{-i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{r}} V(\mathbf{r})$$

is nothing more than the Fourier transform of the potential (where L is the linear size of the sample, so the $\sqrt{L^3}$ terms just normalize the wavefunctions).

Note that these above expressions are true whether or not the sample is a periodic crystal. However, if the sample is periodic the matrix element is zero unless $\mathbf{k} - \mathbf{k}'$ is a reciprocal lattice vector! To see this is true, let us write positions $\mathbf{r} = \mathbf{R} + \mathbf{x}$ where \mathbf{R} is a lattice vector position and \mathbf{x} is a position within the unit cell

$$\langle \mathbf{k}' | V | \mathbf{k} \rangle = \frac{1}{L^3} \int d\mathbf{r} e^{-i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{r}} V(\mathbf{r}) = \frac{1}{L^3} \sum_{\mathbf{R}} \int_{unit-cell} d\mathbf{x} e^{-i(\mathbf{k}' - \mathbf{k}) \cdot (\mathbf{x} + \mathbf{R})} V(\mathbf{x} + \mathbf{R})$$

Now since the potential is assumed periodic, we have $V(\mathbf{x} + \mathbf{R}) = V(\mathbf{x})$, so this can be rewritten as

$$\langle \mathbf{k}' | V | \mathbf{k} \rangle = \frac{1}{L^3} \left[\sum_{\mathbf{R}} e^{-i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{R}} \right] \left[\int_{unit-cell} d\mathbf{x} e^{-i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{x}} V(\mathbf{x}) \right] \quad (13.1)$$

As we discussed in section 12.1.3 above, the first term in brackets must vanish unless $\mathbf{k}' - \mathbf{k}$ is a reciprocal lattice vector³. This condition,

$$\mathbf{k}' - \mathbf{k} = \mathbf{G} \quad (13.2)$$

is known as the *Laue equation* (or *Laue condition*)^{4,5}. This condition is precisely the statement of *the conservation of crystal momentum*.⁶ Note also that when the waves leave the crystal, they

²Fermi's golden rule should be familiar to you from quantum mechanics. Interestingly, Fermi's golden rule was actually discovered by Dirac, giving us yet another example where something is named after Fermi when Dirac really should have credit as well, or even instead. See also footnote 6 in section 4.1.

³Also we discussed that this first term in brackets diverges if $\mathbf{k}' - \mathbf{k}$ is a reciprocal lattice vector. This divergence is not a problem here because it gives just the number of unit cells and is canceled by the $1/L^3$ normalization factor leaving a factor of the inverse volume of the unit cell.

⁴Max von Laue won the Nobel prize for his work on X-ray scattering from crystals in 1914. Although von Laue never left Germany during the second world war, he remained openly opposed to the Nazi government. During the war he hid his gold Nobel medal at the Niels Bohr Institute in Denmark to prevent the Nazis from taking it. Had he been caught doing this, he may have been jailed or worse, since shipping gold out of Nazi Germany was considered a serious offense. After the occupation of Denmark in April 1940, George de Hevesy (a Nobel laureate in chemistry) decided to dissolve the medal in the solvent aqua regia to remove the evidence. He left the solution on a shelf in his lab. Although the Nazis occupied Bohr's institute and searched it very carefully, they did not find anything. After the war, the gold was recovered from solution and the Nobel Foundation presented Laue with a new medal made from the same gold.

⁵The reason this is called "Laue condition" rather than "von Laue" condition is because he was born Max Laue. In 1913 his father was elevated to the nobility and his family added the "von".

⁶Real momentum is conserved since the crystal itself absorbs any missing momentum. In this case, the center of mass of the crystal has absorbed momentum $\hbar(\mathbf{k}' - \mathbf{k})$. See the comment in footnote 9 in section 8.4.

should have

$$|\mathbf{k}| = |\mathbf{k}'|$$

which is just the conservation of energy, which is enforced by the delta function in Fermi's golden rule. (In section 13.4.2 below we will consider more complicated scattering where energy is not conserved.)

13.1.2 Diffraction Approach

It turns out that this Laue condition is nothing more than the scattering condition associated with a diffraction grating. This description of the scattering from crystals is known as the Bragg Formulation of (x-ray) diffraction⁷.

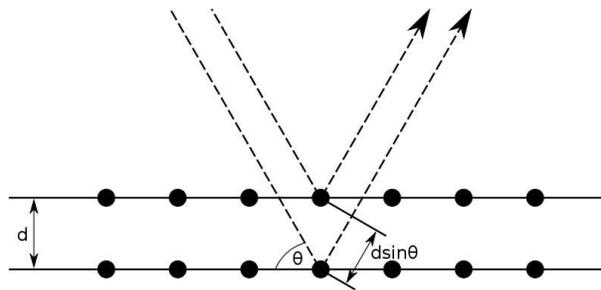


Figure 13.2: Bragg Scattering off of a plane of atoms in a crystal.

Consider the configuration shown in Fig. 13.2. An incoming wave is reflected off of two adjacent layers of atoms separated by a distance d . A few things to note about this diagram. First note that the wave has been deflected by 2θ in this diagram⁸. Secondly, from simple geometry note that the additional distance traveled by the component of the wave that reflects off of the further layer of atoms is

$$\text{extra distance} = 2d \sin \theta.$$

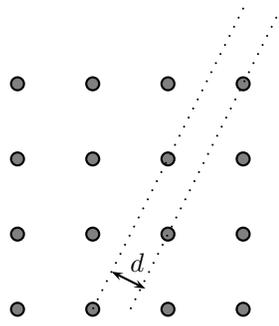
In order to have constructive interference, this extra distance must be equal to an integer number of wavelengths. Thus we derive the Bragg condition for constructive interference, or what is known as *Bragg's law*

$$n\lambda = 2d \sin \theta \tag{13.3}$$

Note that we can have diffraction from any two parallel planes of atoms such as the one shown here

⁷William Henry Bragg and William Lawrence Bragg were a father and son team who won the Nobel prize together in 1915 for their work on X-ray scattering. William Lawrence Bragg was 25 years old when he won the prize, and remains the youngest Nobel laureate ever.

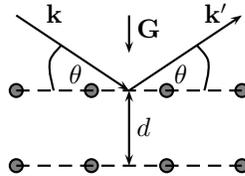
⁸This is a *very* common source of errors on exams. The *total* deflection angle is 2θ .



What we will see next is that this Bragg condition for constructive interference is precisely equivalent to the Laue condition described above.

13.1.3 Equivalence of Laue and Bragg conditions

Consider the following picture (essentially the same as Fig.13.2). Here we have shown the reciprocal lattice vector \mathbf{G} which corresponds to the family of lattice planes. As we discussed in chapter 12 the spacing between lattice planes is $d = 2\pi/|\mathbf{G}|$ (See Eqn. 12.8).



Just from geometry we have

$$\hat{\mathbf{k}} \cdot \hat{\mathbf{G}} = \sin \theta = -\hat{\mathbf{k}}' \cdot \hat{\mathbf{G}}$$

where the hats $\hat{}$ over vectors indicate unit vectors.

Suppose the Laue condition is satisfied. That is, $\mathbf{k} - \mathbf{k}' = \mathbf{G}$ with $|\mathbf{k}| = |\mathbf{k}'| = 2\pi/\lambda$ with λ the wavelength. We can rewrite the Laue equation as

$$\frac{2\pi}{\lambda}(\hat{\mathbf{k}} - \hat{\mathbf{k}}') = \mathbf{G}$$

Now let us dot this equation with $\hat{\mathbf{G}}$ to give

$$\begin{aligned}\hat{\mathbf{G}} \cdot \frac{2\pi}{\lambda}(\hat{\mathbf{k}} - \hat{\mathbf{k}}') &= \hat{\mathbf{G}} \cdot \mathbf{G} \\ \frac{2\pi}{\lambda}(\sin \theta - \sin \theta') &= |\mathbf{G}| \\ \frac{2\pi}{|\mathbf{G}|}(2 \sin \theta) &= \lambda \\ 2d \sin \theta &= \lambda\end{aligned}$$

which is the Bragg condition (in the last step we have used the relation, Eq. 12.8, between \mathbf{G} and d). You may wonder why in this equation we got λ on the right hand side rather than $n\lambda$ as we had in Eq. 13.3. The point here is that there if there is a reciprocal lattice vector \mathbf{G} , then there is also a reciprocal lattice vector $n\mathbf{G}$, and if we did the same calculation with that lattice vector we would get $n\lambda$. In other words, in the $n\lambda$ case we are reflecting off of the spacing nd which necessarily also exists when there is a set of lattice planes with spacing d .

Thus we conclude that the Laue condition and the Bragg condition are equivalent. It is equivalent to say that interference is constructive (as Bragg indicates) or to say that crystal momentum is conserved (as Laue indicates).

13.2 Scattering Amplitudes

If the Laue condition is satisfied, we would now like to ask how much scattering we actually get. Recall in section 13.1.1 we started with Fermi's golden rule

$$\Gamma(\mathbf{k}', \mathbf{k}) = \frac{2\pi}{\hbar} |\langle \mathbf{k}' | V | \mathbf{k} \rangle|^2 \delta(E_{\mathbf{k}'} - E_{\mathbf{k}})$$

and we found out that if V is a periodic function, then the matrix element is given by (See Eq. 13.1)

$$\langle \mathbf{k}' | V | \mathbf{k} \rangle = \left[\frac{1}{L^3} \sum_{\mathbf{R}} e^{-i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{R}} \right] \left[\int_{\text{unit-cell}} d\mathbf{x} e^{-i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{x}} V(\mathbf{x}) \right] \quad (13.4)$$

The first factor in brackets gives zero unless the Laue condition is satisfied, in which case it gives a constant (due to the $1/L^3$ out front, this is now a nondivergent constant). The second term in brackets is known as the structure factor (compare to Eq. 12.6)

$$S(\mathbf{G}) = \int_{\text{unit-cell}} d\mathbf{x} e^{i\mathbf{G} \cdot \mathbf{x}} V(\mathbf{x}) \quad (13.5)$$

where we have used \mathbf{G} for $(\mathbf{k} - \mathbf{k}')$ since this must be a reciprocal lattice vector or the first term in brackets vanishes.

Frequently, one writes

$$I_{(hkl)} \propto |S_{(hkl)}|^2 \quad (13.6)$$

which is shorthand for saying that $I_{(hkl)}$, the intensity of scattering off of the lattice planes defined by the reciprocal lattice vector (hkl) , is proportional to the square of the structure factor at this reciprocal lattice vector. Sometimes a delta-function is also written explicitly to indicate that the wavevector difference $(\mathbf{k}' - \mathbf{k})$ must be a reciprocal lattice vector.

We now turn to examine this structure factor more closely for our main two types of scattering probes – neutrons⁹ and x-rays.

Neutrons

Since neutrons are uncharged, they scatter almost exclusively from nuclei (rather than electrons) via the nuclear forces. As a result, the scattering potential is extremely short ranged, and can be approximated as a delta-function. We thus have

$$V(\mathbf{x}) = \sum_{\text{atom } j \text{ in unit cell}} f_j \delta(\mathbf{x} - \mathbf{x}_j)$$

where \mathbf{x}_j is the position of the j^{th} atom in the unit cell. Here, f_j is known as the *form factor* or *atomic form factor*, and represents the strength of scattering from that particular nucleus. In fact, for the case of neutrons this quantity is proportional to the so-called “nuclear scattering-length” b_j . Thus for neutrons we frequently write

$$V(\mathbf{x}) \sim \sum_{\text{atom } j \text{ in unit cell}} b_j \delta(\mathbf{x} - \mathbf{x}_j)$$

Plugging this expression into Eq. 13.5 above, we obtain

$$S(\mathbf{G}) \sim \sum_{\text{atom } j \text{ in unit cell}} b_j e^{i\mathbf{G}\cdot\mathbf{x}_j} \quad (13.7)$$

X-rays

X-rays scatter from the electrons in a system¹⁰. As a result, one can take $V(\mathbf{x})$ to be proportional to the electron density. We can thus approximate

$$V(\mathbf{x}) \sim \sum_{\text{atom } j \text{ in unit cell}} Z_j g_j(\mathbf{x} - \mathbf{x}_j)$$

where Z_j is the atomic number of atom j (i.e., its number of electrons) and g_j is a somewhat short-ranged function (i.e., it has a few angstroms range — roughly the size of an atom). Taking the Fourier transform, we obtain

$$S(\mathbf{G}) \sim \sum_{\text{atom } j \text{ in unit cell}} f_j(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{x}_j} \quad (13.8)$$

where f_j , the form factor, is roughly proportional to Z_j , but has some dependence on the magnitude of the reciprocal lattice vector \mathbf{G} as well. Frequently, however, we approximate f_j to be independent of \mathbf{G} (which would be true if g were extremely short ranged), although this is not strictly correct.

⁹Brockhouse and Schull were awarded the Nobel prize for pioneering the use of neutron scattering experiments for understanding properties of materials. Schull’s initial development of this technique began around 1946, just after the second world war, when the US atomic energy program made neutrons suddenly available. The Nobel prize was awarded in 1994, making this the longest time-lag ever between a discovery and the awarding of the prize.

¹⁰The coupling of photons to matter is via the usual minimal coupling $(\mathbf{p} + e\mathbf{A})^2/(2m)$. The denominator m , which is much larger for nuclei than for electrons, is why the nuclei are not important.

Aside: As noted above, $f_j(\mathbf{G})$ is just the Fourier transform of the scattering potential for atom j . This scattering potential is proportional to the electron density. Taking the density to be a delta function results in f_j being a constant. Taking the slightly less crude approximation that the density is constant inside a sphere of radius r_0 and zero outside of this radius will result in a Fourier transform

$$f_j(\mathbf{G}) \sim 3Z_j \left(\frac{\sin(x) - x \cos(x)}{x^3} \right) \quad (13.9)$$

with $x = |\mathbf{G}r_0|$ (try showing this!). If the scattering angle is sufficiently small (i.e., \mathbf{G} is small compared to $1/r_0$), the right hand side is roughly Z_j with no strong dependence on \mathbf{G} .

Comparison of Neutrons and X-rays

- For X-rays since $f_j \sim Z_j$ the x-rays scatter very strongly from heavy atoms, and hardly at all from light atoms. This makes it very difficult to “see” light atoms like hydrogen in a solid. Further it is hard to distinguish atoms that are very close to each other in their atomic number (since they scatter almost the same amount). Also f_j is slightly dependent on the scattering angle.
- In comparison the nuclear scattering length b_j varies rather erratically with atomic number (it can even be negative). In particular, hydrogen scatters fairly well, so it is easy to see. Further, one can usually distinguish atoms with similar atomic numbers rather easily.
- For neutrons, the scattering really is very short ranged, so the form factor really is proportional to the scattering length b_j independent of \mathbf{G} . For X-rays there is a dependence of \mathbf{G} that complicates matters.
- Neutrons also have spin. Because of this they can detect whether various electrons in the unit cell have their spins pointing up or down. The scattering of the neutrons from the electrons is much weaker than the scattering from the nuclei, but is still observable. We will return to this situations where the spin of the electron is spatially ordered in section 19.1.2 below.

Simple Example

Generally, as mentioned above, we write the intensity of scattering as

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

Assuming we have orthogonal primitive lattice vectors, we can then generally write

$$S_{(hkl)} = \sum_{\text{atom } j \text{ in unit cell}} f_j e^{2\pi i(hx_j + ky_j + lz_j)} \quad (13.10)$$

where $[x_j, y_j, z_j]$ are the coordinates of atom j within the unit cell, in units of the three primitive lattice vectors.

Example 1: Caesium Chloride: Let us now consider the simple example of CsCl, whose unit cell is shown in Fig. 13.3. This system can be described as simple cubic with a basis given by¹¹

¹¹Do not make the mistake of calling this a lattice! Bcc is a lattice where all points must be the same.

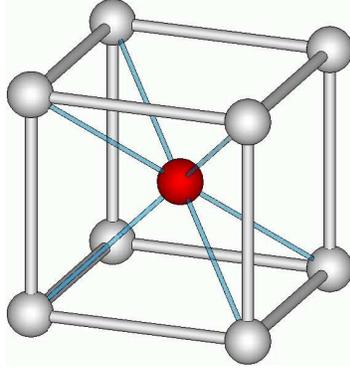


Figure 13.3: Cesium Chloride Unit Cell. Cs is white corner atoms, Cl is red central atom. This is simple cubic with a basis. Note that bcc Cs can be thought of as just replacing the Cl with another Cs atom.

Basis for CsCl	
Cs	Position= $[0, 0, 0]$
Cl	Position= $[a/2, a/2, a/2]$

Thus the structure factor is given by

$$\begin{aligned} S_{(hkl)} &= f_{Cs} + f_{Cl} e^{2\pi i(h,k,l) \cdot [1/2, 1/2, 1/2]} \\ &= f_{Cs} + f_{Cl} (-1)^{h+k+l} \end{aligned}$$

with the f 's being the appropriate form factors for the corresponding atoms. Recall that the scattered wave intensity is $I_{(hkl)} \sim |S_{(hkl)}|^2$.

13.2.1 Systematic Absences and More Examples

Example 2: Caesium bcc: Let us now consider instead a pure Cs crystal. In this case the crystal is bcc. We can think of this as simply replacing the Cl in CsCl with another Cs atom. Analogously we think of the bcc lattice as a simple cubic lattice with exactly the same basis, which we now write as

Basis for Cs bcc	
Cs	Position= $[0, 0, 0]$
Cs	Position= $[a/2, a/2, a/2]$

Now the structure factor is given by

$$\begin{aligned} S_{(hkl)} &= f_{Cs} + f_{Cs} e^{2\pi i(h,k,l) \cdot [1/2, 1/2, 1/2]} \\ &= f_{Cs} [1 + (-1)^{h+k+l}] \end{aligned}$$

Crucially, note that the structure factor, and therefore the scattering intensity vanishes for $h+k+l$ being any odd integer! This phenomenon is known as a *systematic absence*.

To understand why this absence occurs, consider the simple case of the (100) family of planes (See Fig. 12.1). This is simply a family of planes along the crystal axes with spacing a . You might expect a wave of wavelength $2\pi/a$ oriented perpendicular to these planes to scatter constructively. However, if we are considering a bcc lattice, then there are additional planes of atoms half-way between the (100) planes which then cause perfect *destructive* interference. We refer back to the **Important Complication** mentioned in section 12.1.5. As mentioned there, the plane spacing for the bcc lattice in this case is not $2\pi/|\mathbf{G}_{(100)}|$ but is rather $2\pi/|\mathbf{G}_{(200)}|$. In fact in general, for a bcc lattice the plane spacing for any family of lattice planes is $2\pi/|\mathbf{G}_{(hkl)}|$ where $h + k + l$ is always even. This is what causes the selection rule.

Example 3: Copper fcc Quite similarly there are systematic absences for scattering from fcc crystals as well. Recall from Eq. 11.5 that the fcc crystal can be thought of as a simple cubic lattice with a basis given by the points $[0, 0, 0]$, $[1/2, 1/2, 0]$, $[1/2, 0, 1/2]$, and $[0, 1/2, 1/2]$ in units of the cubic lattice constant. As a result the structure factor of fcc copper is given by (plugging into Eq. 13.10)

$$S_{(hkl)} = f_{Cu} \left[1 + e^{i\pi(h+k)} + e^{i\pi(h+l)} + e^{i\pi(k+l)} \right] \quad (13.11)$$

It is easily shown that this expression vanishes unless h, k and l are either all odd or all even.

Summary of Systematic Absences

Systematic Absences of Scattering	
Simple Cubic	all h, k, l allowed
bcc	$h + k + l$ must be even
fcc	h, k, l must be all odd or all even

Systematic absences are sometimes known as *selection rules*.

It is very important to note that these absences, or selection rules, occur for any structure with the given Bravais lattice type. Even if a material is bcc with a basis of five different atoms per primitive unit cell, it will still show the same systematic absences as the bcc lattice we considered above with a single atom per primitive unit cell. To see why this is true we consider yet another example

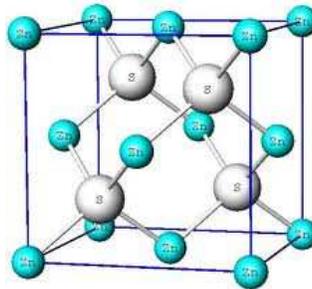


Figure 13.4: Zinc Sulfide Conventional Unit Cell. This is fcc with a basis given by a Zn atom at $[0, 0, 0]$ and a S atom at $[1/4, 1/4, 1/4]$.

Example 4: Zinc Sulfide = fcc with a basis: As shown in Fig 13.4, the Zinc Sulfide

crystal is a an fcc lattice with a basis given by a Zn atom at $[0, 0, 0]$ and an S atom at $[1/4, 1/4, 1/4]$ (this is known as a zincblende structure). If we consider the fcc lattice to itself be a cubic lattice with basis given by the points $[0, 0, 0]$, $[1/2, 1/2, 0]$, $[1/2, 0, 1/2]$, and $[0, 1/2, 1/2]$, we then have the 8 atoms in the conventional unit cell having positions given by the combination of the two bases, i.e.,

		Basis for ZnS			
Zn	Positions=	$[0, 0, 0]$,	$[1/2, 1/2, 0]$,	$[1/2, 0, 1/2]$,	and $[0, 1/2, 1/2]$
S	Positions=	$[1/4, 1/4, 1/4]$,	$[3/4, 3/4, 1/4]$,	$[3/4, 1/4, 3/4]$,	and $[1/4, 3/4, 3/4]$

The structure factor for ZnS is thus given by

$$S_{(hkl)} = f_{Zn} \left[1 + e^{2\pi i(hkl) \cdot [1/2, 1/2, 0]} + \dots \right] + f_S \left[e^{2\pi i(hkl) \cdot [1/4, 1/4, 1/4]} + e^{2\pi i(hkl) \cdot [3/4, 3/4, 1/4]} + \dots \right]$$

This combination of 8 terms can be factored to give

$$S_{(hkl)} = \left[1 + e^{i\pi(h+k)} + e^{i\pi(h+l)} + e^{i\pi(k+l)} \right] \left[f_{Zn} + f_S e^{i(\pi/2)(h+k+l)} \right] \quad (13.12)$$

The first term in brackets is precisely the same as the term we found for the fcc crystal in Eq. 13.11. In particular it has the same systematic absences that it vanishes unless h, k and l are either all even or all odd. The second term gives additional absences associated specifically with the ZnS structure.

Since the positions of the atoms are the positions of the underlying lattice plus the vectors in the basis, it is easy to see that the structure factor of a crystal system with a basis will always factorize into a piece which comes from the underlying lattice structure times a piece corresponding to the basis. Generalizing Eq. 13.12 we can write

$$S_{(hkl)} = S_{(hkl)}^{Lattice} \times S_{(hkl)}^{basis} \quad (13.13)$$

(where, to be precise, the form factors only occur in the latter term).

13.3 Methods of Scattering Experiments

There are many methods of performing scattering experiments. In principle they are all similar — one sends in a probe wave of known wavelength (an X-ray, for example) and measures the angles at which it diffracts when it comes out. Then using Bragg's laws (or the Laue equation) one can deduce the spacings of the lattice planes in the system.

13.3.1 Advanced Methods (interesting and useful but you probably won't be tested on this)

Laue Method

Conceptually, perhaps the simplest method is to take a large single crystal of the material in question — fire waves at it (X-rays, say) from one direction, and measure the direction of the outgoing waves. However, given a single direction of incoming wave, it is unlikely that you precisely achieve the diffraction condition (the Bragg condition) for any given set of lattice planes. In order to get more data, one can then vary the wavelength of the incoming wave. This allows one to achieve the Bragg condition, at least at some wavelength.

Rotating Crystal Method

A similar technique is to rotate the crystal continuously so that at some angle of the crystal with respect to the incoming waves, one achieves the Bragg condition and measures an outgoing diffracted wave.

Both of these methods are used. However, there is an important reason that they are sometimes impossible. Frequently it is not possible to obtain a single crystal of a material. Growing large crystals (such as the beautiful ones shown in Fig. 6) can be an enormous challenge¹² In the case of neutron scattering, the problem is even more acute since one typically needs fairly large single crystals compared to x-rays.

13.3.2 Powder Diffraction (you will almost certainly be tested on this!)

Powder diffraction, or the *Debye-Scherrer method*¹³ is the use of wave scattering on a sample which is not single crystalline, but is powdered. In this case, the incoming wave can scatter off of any one of many small crystallites which may be oriented in any possible direction. In spirit this technique is similar to the rotating crystal method in that there is always some angle at which a crystal can be oriented to diffract the incoming wave. A figure of the Debye-Scherrer setup is shown in Fig. 13.5. Using Bragg's law, given the wavelength of the incoming wave, we can deduce the possible spacings between lattice planes.

A Fully Worked Example. Study this!

Because this type of problem has historically ended up on exams essentially every year, and because it is hard to find references that explain how to solve these problems, I am going to work a powder-diffraction problem in detail here. As far as I can tell, they will only ever ask you about cubic lattices (simple cubic, fcc, and bcc).

Before presenting the problem and solving it, however, it is useful to write down a table of possible lattice planes and the selection rules that can occur for the smallest reciprocal lattice vectors

¹²For example, high-temperature superconducting materials were discovered in 1986 (and resulted in a Nobel prize the next year!). Despite a concerted world-wide effort, good single crystals of these materials were not available for 5 to 10 years.

¹³Debye is the same guy from the specific heat of solids. Paul Scherrer was Swiss but worked in Germany during the second world war, where he passed information to the famous American spy (and baseball player), Moe Berg, who had been given orders to find and shoot Heisenberg if he felt that the Germans were close to developing a bomb.

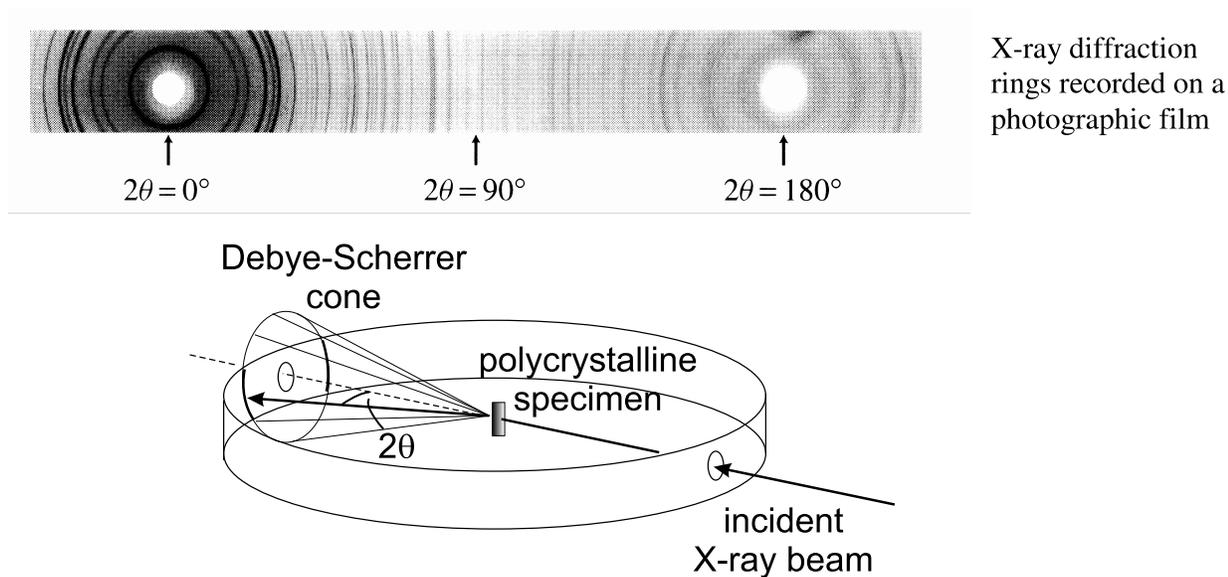


Figure 13.5: Debye-Scherrer Powder Diffraction.

Lattice Plane Selection Rules						
$\{hkl\}$	$N = h^2 + k^2 + l^2$	Multiplicity	cubic	bcc	fcc	
100	1	6	✓			
110	2	12	✓	✓		
111	3	8	✓		✓	
200	4	6	✓	✓	✓	
210	5	24	✓			
211	6	24	✓	✓		
220	8	12	✓	✓	✓	
221	9	24	✓			
300	9	6	✓			
310	10	24	✓	✓		
311	11	24	✓		✓	
222	12	8	✓	✓	✓	
⋮	⋮					

The selection rules are exactly those listed above: simple cubic allows scattering from any plane, bcc must have $h + k + l$ be even, and fcc must have h, k, l either all odd or all even.

We have added a column N which is the square magnitude of the reciprocal lattice vector.

We have also added an additional column labeled “multiplicity”. This quantity is important

for figuring out the amplitude of scattering. The point here is that the (100) planes have some particular spacing but there are 5 other families of planes with the same spacing: (010), (001), ($\bar{1}00$), ($0\bar{1}0$), ($00\bar{1}$). (Because we mean all of these possible families of lattice planes, we use the notation $\{hkl\}$ introduced at the end of section 12.1.5). In the powder diffraction method, the crystal orientations are random, and here there would be 6 possible equivalent orientations of a crystal which will present the right angle for scattering from one of these planes, so there will be scattering intensity which is 6 times as large as we would otherwise calculate — this is known as the multiplicity factor. For the case of the 111 family, we would instead find 8 possible equivalent planes: (111), (11 $\bar{1}$), ($\bar{1}11$), (1 $\bar{1}\bar{1}$), ($\bar{1}\bar{1}1$), (1 $\bar{1}\bar{1}$), ($\bar{1}\bar{1}\bar{1}$), ($\bar{1}\bar{1}\bar{1}$). Thus, we should replace Eq. 13.6 with

$$I_{\{hkl\}} \propto M_{\{hkl\}} |S_{\{hkl\}}|^2 \quad (13.14)$$

where M is the multiplicity factor.

Calculating this intensity is straightforward for neutron scattering, but is much harder for x-ray scattering because the form factor for X-rays depends on \mathbf{G} . I.e, since in Eq. 13.7 the form factor (or scattering length b_j) is a constant independent of \mathbf{G} , it is easy to calculate the expected amplitudes of scattering based only on these constants. For the case of X-rays you need to know the functional forms of $f_j(\mathbf{G})$. At some *very* crude level of approximation it is a constant. More precisely we see in Eq. 13.9 that it is constant for small scattering angle but can vary quite a bit for large scattering angle.

Even if one knows the detailed functional form of $f_j(\mathbf{G})$, experimentally observed scattering intensities are never quite of the form predicted by Eq. 13.14. There can be several sources of corrections¹⁴ that modify this result (these corrections are usually swept under the rug in elementary introductions to scattering, but you should at least be aware that they exist). Perhaps the most significant corrections¹⁵ are known as Lorentz corrections or Lorentz-Polarization corrections. These terms, which depend on the detailed geometry of the experiment, give various prefactors (involving terms like $\cos \theta$ for example) which are smooth as a function of θ .

The Example

Consider the powder diffraction data from PrO_2 shown in Fig. 13.6. (Exactly this data was presented in the 2009 Exam, and we were told that the lattice is some type of cubic lattice. As we will see below there were several small, but important, errors in the question!) Given the wavelength .123 nm, we first would like to figure out the type of lattice and the lattice constant.

Note that the full deflection angle is 2θ . We will want to use Bragg's law and the expression for the spacing between planes

$$d_{(hkl)} = \frac{\lambda}{2 \sin \theta} = \frac{a}{\sqrt{h^2 + k^2 + l^2}}$$

where we have also used the expression Eq. 12.12 for the spacing between planes in a cubic lattice given the lattice constant a . Note that this then gives us

$$a^2/d^2 = h^2 + k^2 + l^2 = N$$

¹⁴Many of these corrections were first worked out by Charles Galton Darwin, the grandson of Charles Robert Darwin, the brilliant naturalist and proponent of evolution. The younger Charles was a terrific scientist in his own right. Later in life his focus turned to ideas of eugenics, predicting that the human race would eventually fail as we continue to breed unfavorable traits. (His interest in eugenics is not surprising considering that the acknowledged father of eugenics, Francis Galton, was also part of the same family.)

¹⁵Another important correction is due to the thermal vibrations of the crystal. Using Debye's theory of vibration, Ivar Waller derived what is now known as the Debye-Waller factor that accounts for the thermal smearing of Bragg peaks.

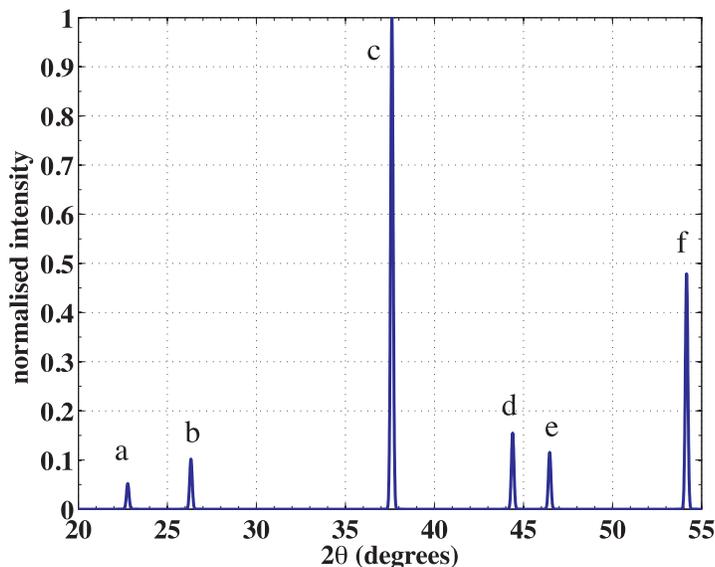


Figure 13.6: Powder Diffraction of Neutrons from PrO_2 . The wavelength of the neutron beam is $\lambda = .123$ nm. (One should assume that Lorentz corrections have been removed from the displayed intensities.)

which is what we have labeled N in the above table of selection rules. We now make a table. In the first two columns we just read the angles off of the given graph. *You should try to make the measurements of the angle from the data as carefully as possible. It makes the analysis much easier if you measure the angles right!*

peak	2θ	$d = \lambda/(2 \sin \theta)$	d_a^2/d^2	$3d_a^2/d^2$	$N = h^2 + k^2 + l^2$	$\{hkl\}$	$a = d\sqrt{h^2 + k^2 + l^2}$
a	22.7°	0.313 nm	1	3	3	111	.542 nm
b	26.3°	0.270 nm	1.33	3.99	4	200	.540 nm
c	37.7°	0.190 nm	2.69	8.07	8	220	.537 nm
d	44.3°	0.163 nm	3.67	11.01	11	311	.541 nm
e	46.2°	0.157 nm	3.97	11.91	12	222	.544 nm
f	54.2°	0.135 nm	5.35	16.05	16	400	.540 nm

In the third column of the table we calculate the distance between lattice planes for the given diffraction peak using Bragg's law. In the fourth column we have calculated the squared ratio of the lattice spacing d for the given peak to the lattice spacing for the first peak (labeled a) as a reference. We then realize that these ratios are pretty close to whole numbers divided by three, so we try multiplying each of these quantities by 3 in the next column. If we round these numbers to integers (given in the next column), we produce precisely the values of $N = h^2 + k^2 + l^2$ expected for the fcc lattice (According to the above selection rules we must have h, k, l all even or all odd). The final column calculates the lattice constant from the given diffraction angle. Averaging these numbers gives us a measurement of the lattice constant $a = .541 \pm .002$ nm.

The analysis thus far is equivalent to what one would do for X-ray scattering. However, with neutrons, assuming the scattering length is independent of scattering angle (which is typically a good assumption) we can go a bit further by analyzing the the intensity of the scattering peaks.

In real data often intensities are weighted by the above mentioned Lorentz factors. In Fig. 13.6 these factors have been removed so that we can expect that Eq. 13.14 holds precisely. (One error in the Exam question was that it was not mentioned that these factors have been removed!)

In the problem given on the 2009 Exam, it is given that the basis for this crystal is a Pr atom at position $[0,0,0]$ and O at $[1/4,1/4,1/4]$ and $[1/4,1/4,3/4]$. Thus, the Pr atoms form a fcc lattice and the O's fill in the holes as shown in Fig. 13.7.

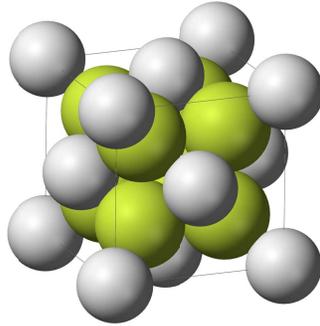


Figure 13.7: The fluorite structure. This is fcc with a basis given by a white atom (Pr) at $[0,0,0]$ and yellow atoms (O) at $[1/4, 1/4, 1/4]$ and $[1/4, 1/4, 3/4]$.

Let us calculate the structure factor for this crystal. Using Eq. 13.13 we have

$$S_{(hkl)} = \left[1 + e^{i\pi(h+k)} + e^{i\pi(h+l)} + e^{i\pi(k+l)} \right] \left[b_{Pr} + b_O \left(e^{i(\pi/2)(h+k+l)} + e^{i(\pi/2)(h+k+3l)} \right) \right]$$

The first term in brackets is the structure factor for the fcc lattice, and it gives 4 for every allowed scattering point (when h, k, l are either all even or all odd). The second term in brackets is the structure factor for the basis.

The scattering intensity of the peaks are then given in terms of this structure factor and the peak multiplicities as shown in Eq. 13.14. We thus can write for all of our measured peaks¹⁶

$$I_{\{hkl\}} = CM_{\{hkl\}} \left| b_{Pr} + b_O \left(e^{i(\pi/2)(h+k+l)} + e^{i(\pi/2)(h+k+3l)} \right) \right|^2$$

where the constant C contains other constant factors (including the factor of 4^2 from the fcc structure factor). Note: We have to be a bit careful here to make sure that the bracketed factor gives the same result for all possible (hkl) included in $\{hkl\}$, but in fact it does. Thus we can compile another table showing the predicted relative intensities of the peaks.

¹⁶Again assuming that smooth Lorentz correction terms have been removed from our data so that Eq. 13.14 is accurate.

Scattering Intensity			
peak	{hkl}	$I_{\{hkl\}}/C \propto M S ^2$	Measured Intensity
a	111	$8b_{Pr}^2$	0.05
b	200	$6[b_{Pr} - 2b_O]^2$	0.1
c	220	$12[b_{Pr} + 2b_O]^2$	1.0
d	311	$24b_{Pr}^2$	0.15
e	222	$8[b_{Pr} - 2b_O]^2$	0.1
f	400	$6[b_{Pr} + 2b_O]^2$	0.5

where the final column are the intensities measured from the data in Fig. 13.6.

From the analytic expressions in the third column we can immediately predict that we should have

$$I_d = 3I_a \quad I_c = 2I_f \quad I_e = \frac{4}{3}I_b$$

Examining the fourth column of this table, it is clear that the first two of these equations are properly satisfied. However the final equation does not appear to be correct. This points to some error in constructing the plot. Thus we suspect some problem in either I_e or I_b . Either I_e is too small or I_b is too large¹⁷.

To further home in on this problem with the data, we can look at the ratio I_c/I_a which in the measured data has a value of about 20. Thus we have

$$\frac{I_c}{I_a} = \frac{12[b_{Pr} + 2b_O]^2}{8b_{Pr}^2} = 20$$

with some algebra this can be reduced to a quadratic equation with two roots, resulting in

$$b_{Pr} = -.43b_O \quad \text{or} \quad .75b_O \quad (13.15)$$

Let us suppose now that our measurement of I_b is correct. In this case we have

$$\frac{I_b}{I_a} = \frac{6[b_{Pr} - 2b_O]^2}{8b_{Pr}^2} = 2$$

which we can solve to give

$$b_{Pr} = .76b_O \quad \text{or} \quad -3.1b_O$$

The former solution being reasonably consistent with the above. However, were we to assume I_e were correct, we would have instead

$$\frac{I_e}{I_a} = \frac{8[b_{Pr} - 2b_O]^2}{8b_{Pr}^2} = 2$$

we would obtain

$$b_{Pr} = .83b_O \quad \text{or} \quad -4.8b_O$$

which appears inconsistent with Eq. 13.15. We thus conclude that the measured intensity of I_e given in Fig. 13.6 is actually incorrect, and should be larger by about a factor of 4/3. (This is the second error in the exam question.) We can then conclude that Having now corrected this error, we note that we have now used this neutron data to experimentally determine the ratio of the nuclear scattering lengths

$$b_{Pr}/b_O \approx .75$$

¹⁷Another possibility is that the form factor is not precisely independent of scattering angle, as is the case for X-ray scattering. However, the fact that all the peaks are consistent but for this one peak suggests a transcription error.

13.4 Still more about scattering

Scattering experiments such as those discussed here are *the* method for determining the microscopic structures of materials. One can use these methods (and extensions thereof) to sort out even very complicated atomic structures such as those of biological molecules.

Aside: In addition to the obvious work of von Laue and Bragg that initiated the field of X-ray diffraction (and Brockhouse and Schull for neutrons) there have been about half a dozen Nobel prizes that have relied on, or further developed these techniques. In 1962 a chemistry Nobel prize was awarded to Perutz and Kendrew for using X-rays to determine the structure of the biological proteins hemoglobin and myoglobin. The same year, Watson and Crick were awarded the prize in Biology for determining the structure of DNA — which they did with the help of X-ray diffraction data taken by Rosalind Franklin¹⁸. Two years later in 1964, Dorothy Hodgkin¹⁹ won the prize for determination of the structure of penicillin and other biological molecules. Further Nobels were given in chemistry for determining the structure of Boranes (Lipscomb, 1976) and for the structure of photosynthetic proteins (Deisenhofer, Huber, Michel 1988).

13.4.1 Variant: Scattering in Liquids and Amorphous Solids

A material need not be crystalline to scatter waves. However, for amorphous solids or liquids, instead of having delta-function peaks in the structure factor at reciprocal lattice vectors (as in Fig. 13.6), the structure factor (which is again defined as the Fourier transform of the density) will have smooth behavior — with incipient peaks corresponding to $2\pi/d$ where d is roughly the typical distance between atoms. An example of a measured structure factor in liquid Al is shown in Fig. 13.8. As the material gets close to its freezing point, the peaks in the structure factor will get more pronounced, becoming more like the structure of a solid where the peaks are delta functions.

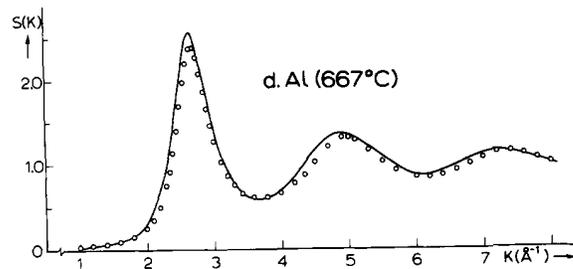


Figure 13.8: The structure factor of liquid Aluminum

¹⁸There remains quite a controversy over the fact that Watson and Crick, at a critical juncture, were shown Franklin's data without her knowledge! Franklin may have won the prize in addition to Watson and Crick and thereby received a bit more of the appropriate credit, but she tragically died of cancer at age 37 in 1958, two years before the prize was awarded.

¹⁹Dorothy Hodgkin was a student and later a fellow at [Somerville College, Oxford](#). Yay!

13.4.2 Variant: Inelastic Scattering

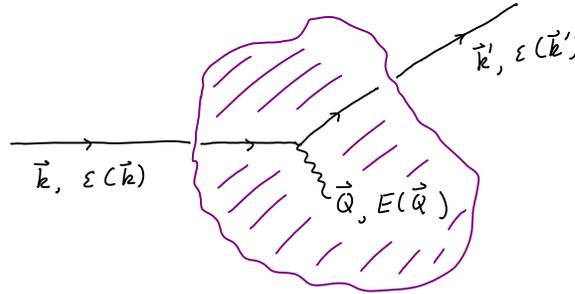


Figure 13.9: Inelastic scattering. Energy and crystal momentum must be conserved.

It is also possible to perform scattering experiments which are inelastic. Here, “inelastic” means that energy of the incoming wave is left behind in the sample, and the energy of the outgoing wave is lower. The general process is shown in Fig. 13.9. A wave is incident on the crystal with momentum \mathbf{k} and energy $\epsilon(\mathbf{k})$ (For neutrons the energy would be $\hbar^2\mathbf{k}^2/(2m)$ whereas for photons the energy would be $\hbar c|\mathbf{k}|$). This wave transfers some of its energy and momentum to some internal excitation mode of the material — such as a phonon, or a spin or electronic excitation quanta. One then measures the outgoing energy and momentum of the wave. Since energy and crystal momentum must be conserved, one has

$$\begin{aligned}\mathbf{Q} &= \mathbf{k}' - \mathbf{k} + \mathbf{G} \\ E(\mathbf{Q}) &= \epsilon(\mathbf{k}') - \epsilon(\mathbf{k})\end{aligned}$$

thus allowing one to determine the dispersion relation of the internal excitation (i.e., the relationship between \mathbf{Q} and $E(\mathbf{Q})$). This technique is extremely useful for determining phonon dispersions experimentally. In practice, the technique is much more useful with neutrons than with X-rays. The reason for this is, because the speed of light is so large, (and $E = \hbar c|\mathbf{k}|$) the energy differences that one obtains are enormous except for a tiny small range of \mathbf{k}' for each \mathbf{k} . Since there is a maximum energy for a phonon, the X-rays therefore have a tiny total cross section for exciting phonons. A second reason that this technique is difficult for X-rays is because it is much harder to build an X-ray detector that determines energy than it is for neutrons.

13.4.3 Experimental Apparatus

Perhaps the most interesting piece of this kind of experiments is the question of how one actually produces and measures the waves in questions.

Since at the end of the day, one ends up counting photons or neutrons, brighter sources (higher flux of probe particles) are always better — as it allows one to do experiments quicker, and allows one to reduce noise (since counting error on N counts is proportional to \sqrt{N} , meaning a fractional error that drops as $1/\sqrt{N}$). Further, with a brighter source, one can examine smaller samples more easily.

X-rays: Even small laboratories can have X-ray sources that can do very useful crystallography. A typical source accelerates electrons electrically (with 10s of keV) and smashes them into a metal target. X-rays with a discrete spectrum of energies are produced when an electron is knocked out of a low atomic orbital and an electron in a higher orbital drops down to re-fill the hole (this is known as X-ray fluorescence). Also a continuous Bremsstrahlung spectrum is produced by electrons coming near the charged nuclei, but for monochromatic diffraction experiments, this is less useful. (One wavelength from a spectrum can be selected — using diffraction from a known crystal!).

Much higher brightness X-ray sources are provided by huge (and hugely expensive) facilities known as synchrotron light sources — where particles (usually electrons) are accelerated around enormous loops (at energies in the GeV range). Then using magnets these electrons are rapidly accelerated around corners which makes them emit X-rays extremely brightly and in a highly columnated fashion.

Detection of X-rays can be done with photographic films (the old style) but is now more frequently done with more sensitive semiconductor detectors.

Neutrons: Although it is possible to generate neutrons in a small lab, the flux of these devices is extremely small and neutron scattering experiments are always done in large neutron source facilities. Although the first neutron sources simply used the byproduct neutrons from nuclear reactors, more modern facilities now use a technique called *spallation* where protons are accelerated into a target and neutrons are emitted. As with X-rays, neutrons can be monochromated (made into a single wavelength), by diffracting them from a known crystal. Another technique is to use time-of-flight. Since more energetic neutrons move faster, one can send a pulse of poly-chromatic neutrons and select only those that arrive at a certain time in order to obtain mono-chromatic neutrons. On the detection side, one can again select for energy very easily. I won't say too much about neutron detection as there are many methods. Needless to say, they all involve interaction with nuclei.

13.5 Summary of Diffraction

- Diffraction of waves from crystals in Laue and Bragg formulations (equivalent to each other).
- The structure factor (the Fourier transform of the scattering potential) in a periodic crystal has sharp peaks at allowed reciprocal lattice vectors for scattering. The scattering intensity is the square of the structure factor.
- There are systematic absences of diffraction peaks depending on the crystal structure (fcc, bcc). Know how to figure these out.
- Know how to analyze a powder diffraction pattern (very common exam question!)

References

It is hard to find references that give enough information about diffraction to suit the Oxford course. These are not bad.

- Kittel, chapter 2
- Ashcroft and Mermin, chapter 6



Figure 13.10: The Rutherford-Appleton Lab in Oxfordshire, UK. On the right, the large circular building is the DIAMOND synchrotron light source. The building on the left is the ISIS spallation neutron facility. This was the world's brightest neutron source on earth until August 2007 when it was surpassed by one in Oak Ridge, US. The next generation source is being built in Sweden and is expected to start operating in 2019. The price tag for construction of this device is over 10^9 euros.

- Dove, chapter 6 (most detailed, with perhaps a bit too much information in places)

In addition, the following have nice, but incomplete discussions.

- Rosenberg, chapter 2.
- Ibach and Luth, chapter 3.
- Burns, chapter 4.

Part VI

Electrons in Solids

Chapter 14

Electrons in a Periodic Potential

In chapters 8 and 9 we discussed the wave nature of phonons in solids, and how crystal momentum is conserved (i.e., momentum is conserved up to reciprocal lattice vector). Further we found that we could describe the entire excitation spectrum within a single Brillouin zone in a reduced zone scheme. We also found in chapter 13 that X-rays and neutrons similarly scatter from solids by conserving crystal momentum. In this chapter we will consider the nature of electron waves in solids and we will find that similarly crystal momentum is conserved and the entire excitation spectrum can be described within a single Brillouin zone using a reduced zone scheme.

We have seen a detailed preview of properties of electrons in periodic systems when we considered the one dimensional tight binding model in chapter 10, so the results of this section will be hardly surprising. However, in the current chapter we will approach the problem from a very different (and complimentary) starting point. Here, we will consider electrons as free-electron waves that are very only very weakly perturbed by the periodic arrangement of atoms in the solid. The tight binding model is exactly the opposite limit where we consider electrons bound strongly to the atoms, and they only weakly hop from one atom to the next.

14.1 Nearly Free Electron Model

We start with completely free electrons whose Hamiltonian is

$$H_0 = \frac{\mathbf{p}^2}{2m}$$

The corresponding energy eigenstates, the plane waves $|\mathbf{k}\rangle$, have eigenenergies

$$\epsilon_0(\mathbf{k}) = \frac{\hbar^2|\mathbf{k}|^2}{2m}$$

We now consider a weak periodic potential perturbation to this Hamiltonian

$$H = H_0 + V(\mathbf{r})$$

with

$$V(\mathbf{r}) = V(\mathbf{r} + \mathbf{R})$$

where \mathbf{R} is any lattice vector. The matrix elements of this potential are then just the Fourier components

$$\langle \mathbf{k}' | V | \mathbf{k} \rangle = \frac{1}{L^3} \int \mathbf{dr} e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}} V(\mathbf{r}) \equiv V_{\mathbf{k}'-\mathbf{k}} \quad (14.1)$$

which is zero unless $\mathbf{k}' - \mathbf{k}$ is a reciprocal lattice vector (See Eq. 13.1). Thus, any plane wave state \mathbf{k} can scatter into another plane wave state \mathbf{k}' only if these two plane waves are separated by a reciprocal lattice vector.

We now apply the rules of perturbation theory. At first order in the perturbation V , we have

$$\epsilon(\mathbf{k}) = \epsilon_0(\mathbf{k}) + \langle \mathbf{k} | V | \mathbf{k} \rangle = \epsilon_0(\mathbf{k}) + V_0$$

which is just an uninteresting constant energy shift to all of the eigenstates. In fact, it is an exact statement (at any order of perturbation theory) that the only effect of V_0 is to shift the energies of all of the eigenstates by this constant¹. Henceforth we will assume that $V_0 = 0$ for simplicity.

At second order in perturbation theory we have

$$\epsilon(\mathbf{k}) = \epsilon_0(\mathbf{k}) + V_0 + \sum_{\mathbf{k}'=\mathbf{k}+\mathbf{G}}' \frac{|\langle \mathbf{k}' | V | \mathbf{k} \rangle|^2}{\epsilon_0(\mathbf{k}) - \epsilon_0(\mathbf{k}')} \quad (14.2)$$

where the $'$ means that the sum is restricted to have $\mathbf{G} \neq \mathbf{0}$. In this sum, however, we have to be careful. It is possible that, for some \mathbf{k}' it happens that $\epsilon_0(\mathbf{k})$ is very close to $\epsilon_0(\mathbf{k}')$ or perhaps they are even equal. In which case the corresponding term of the sum diverges and the perturbation expansion makes no sense. This is what we call a *degenerate* situation and it needs to be handled with degenerate perturbation theory, which we shall consider below.

To see when this degenerate situation happens, we look for solutions of

$$\epsilon_0(\mathbf{k}) = \epsilon_0(\mathbf{k}') \quad (14.3)$$

$$\mathbf{k}' = \mathbf{k} + \mathbf{G} \quad (14.4)$$

First, let us consider the one-dimensional case. Since $\epsilon(k) \sim k^2$, the only possible solutions of Eq. 14.3 is $k' = -k$. This means the two equations are only satisfied for

$$k' = -k = \pi n/a$$

or precisely on the Brillouin zone boundaries (See Fig. 14.1).

In fact, this is quite general even in higher dimensions: given a point \mathbf{k} on a Brillouin zone boundary, there is another point \mathbf{k}' (also on a Brillouin zone boundary) such that Eqs. 14.3 and 14.4 are satisfied (See in particular Fig. 12.5 for example)².

Since Eq. 14.2 is divergent, we need to handle this situation with *degenerate perturbation theory*³. In this approach, one diagonalizes the Hamiltonian within the degenerate space first (and other perturbations can be treated after this). In other words, we take states of the same energy that are connected by the matrix element and treat their mixing exactly.

¹You should be able to show this!

²To see this generally, recall that a Brillouin zone boundary is a perpendicular bisector of the segment between $\mathbf{0}$ and some \mathbf{G} . We can write the given point $\mathbf{k} = \mathbf{G}/2 + \mathbf{k}_\perp$ where $\mathbf{k}_\perp \cdot \mathbf{G} = 0$. Then if we construct the point $\mathbf{k}' = -\mathbf{G}/2 + \mathbf{k}_\perp$, then clearly 14.4 is satisfied, \mathbf{k}' is a perpendicular bisector of the segment between $\mathbf{0}$ and $-\mathbf{G}$ and therefore is on a zone boundary, and $|\mathbf{k}| = |\mathbf{k}'|$ which implies that Eq. 14.3 is satisfied.

³Hopefully you have learned this in your quantum mechanics courses already!

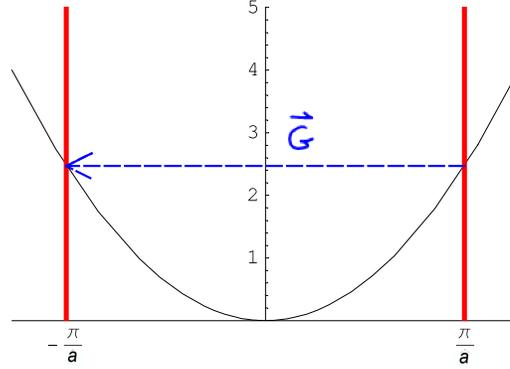


Figure 14.1: Scattering from Brillouin Zone Boundary to Brillouin Zone Boundary. The states at the two zone boundaries are separated by a reciprocal lattice vector \mathbf{G} and have the same energy. This situation leads to a divergence in perturbation theory, Eq. 14.2 because when the two energies match, the denominator is zero.

14.1.1 Degenerate Perturbation Theory

If two plane wave states $|\mathbf{k}\rangle$ and $|\mathbf{k}'\rangle = |\mathbf{k} + \mathbf{G}\rangle$ are of approximately the same energy (meaning that \mathbf{k} and \mathbf{k}' are close to zone boundaries), then we must diagonalize the matrix elements of these states first. We have

$$\begin{aligned} \langle \mathbf{k} | H | \mathbf{k} \rangle &= \epsilon_0(\mathbf{k}) \\ \langle \mathbf{k}' | H | \mathbf{k}' \rangle &= \epsilon_0(\mathbf{k}') = \epsilon_0(\mathbf{k} + \mathbf{G}) \\ \langle \mathbf{k} | H | \mathbf{k}' \rangle &= V_{\mathbf{k}-\mathbf{k}'} = V_{\mathbf{G}}^* \\ \langle \mathbf{k}' | H | \mathbf{k} \rangle &= V_{\mathbf{k}'-\mathbf{k}} = V_{\mathbf{G}} \end{aligned} \quad (14.5)$$

where we have used the definition of $V_{\mathbf{G}}$ from Eq. 14.1, and the fact that $V_{-\mathbf{G}} = V_{\mathbf{G}}^*$ is guaranteed by the fact that $V(\mathbf{r})$ is real.

Now, within this two dimensional space we can write any wavefunction as

$$|\Psi\rangle = \alpha|\mathbf{k}\rangle + \beta|\mathbf{k}'\rangle = \alpha|\mathbf{k}\rangle + \beta|\mathbf{k} + \mathbf{G}\rangle \quad (14.6)$$

Using the variational principle to minimize the energy is equivalent to solving the effective Schroedinger equation⁴

$$\begin{pmatrix} \epsilon_0(\mathbf{k}) & V_{\mathbf{G}}^* \\ V_{\mathbf{G}} & \epsilon_0(\mathbf{k} + \mathbf{G}) \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = E \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \quad (14.7)$$

The secular equation determining E is then

$$\begin{pmatrix} \epsilon_0(\mathbf{k}) - E \\ \epsilon_0(\mathbf{k} + \mathbf{G}) - E \end{pmatrix} - |V_{\mathbf{G}}|^2 = 0 \quad (14.8)$$

(Note that once this degenerate space is diagonalized, one could go back and treat further, nondegenerate, scattering processes, in perturbation theory.)

⁴This should look similar to our 2 by 2 Schroedinger Equation 5.8 above.

Simple Case: \mathbf{k} exactly at the zone boundary

The simplest case we can consider is when \mathbf{k} is precisely on a zone boundary (and therefore $\mathbf{k}' = \mathbf{k} + \mathbf{G}$ is also precisely on a zone boundary). In this case $\epsilon_0(\mathbf{k}) = \epsilon_0(\mathbf{k} + \mathbf{G})$ and our secular equation simplifies to

$$\left(\epsilon_0(\mathbf{k}) - E \right)^2 = |V_{\mathbf{G}}|^2$$

or equivalently

$$E_{\pm} = \epsilon_0(\mathbf{k}) \pm |V_{\mathbf{G}}|$$

Thus we see that a gap opens up at the zone boundary. Whereas both \mathbf{k} and \mathbf{k}' had energy $\epsilon_0(\mathbf{k})$ in the absence of the added potential $V_{\mathbf{G}}$, when the potential is added, the two eigenstates form two linear combinations with energies split by $\pm|V_{\mathbf{G}}|$.

In one dimension

In order to understand this better, let us focus on the one dimensional case. Let us assume we have a potential $V(x) = \tilde{V} \cos(2\pi x/a)$ with $\tilde{V} > 0$. The Brillouin zone boundaries are at $k = \pi/a$ and $k' = -k = -\pi/a$ so that $k' - k = G = -2\pi/a$ and $\epsilon_0(k) = \epsilon_0(k')$.

Examining Eq. 14.7, we discover that the solutions (when $\epsilon_0(k) = \epsilon_0(k')$) are given by $\alpha = \pm\beta$ thus giving the eigenstates

$$|\psi_{\pm}\rangle = \frac{1}{\sqrt{2}} (|k\rangle \pm |k'\rangle) \quad (14.9)$$

corresponding to E_{\pm} respectively. Since we can write the real space version of these $|k\rangle$ wavefunctions as⁵

$$\begin{aligned} |k\rangle &\rightarrow e^{ikx} = e^{ix\pi/a} \\ |k'\rangle &\rightarrow e^{-ik'x} = e^{-ix\pi/a} \end{aligned}$$

we discover that the two eigenstates are given by

$$\begin{aligned} \psi_+ &\sim e^{ix\pi/a} + e^{-ix\pi/a} \propto \cos(x\pi/a) \\ \psi_- &\sim e^{ix\pi/a} - e^{-ix\pi/a} \propto \sin(x\pi/a) \end{aligned}$$

If we then look at the densities $|\psi_{\pm}|^2$ associated with these two wavefunctions (See Fig. 14.2) we see that the higher energy eigenstate ψ_+ has its density concentrated mainly at the maxima of the potential V whereas the lower energy eigenstate ψ_- has its density concentrated mainly at the minima of the potential.

So the general principle is that the periodic potential scatters between the two plane waves \mathbf{k} and $\mathbf{k} + \mathbf{G}$. If the energy of these two plane waves are the same, the mixing between them is strong, and the two plane waves can combine to form one state with higher energy (concentrated on the potential maxima) and one state with lower energy (concentrated on the potential minima).

⁵Formally what we mean here is $\langle x|k\rangle = e^{ikx}/\sqrt{L}$.

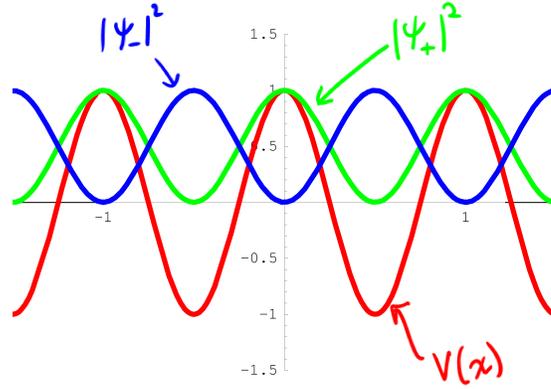


Figure 14.2: Structure of Wavefunctions at the Brillouin Zone Boundary. The higher energy eigenstate ψ_+ has its density concentrated near the maxima of the potential V whereas the lower energy eigenstate has its density concentrated near the minima.

k not quite on a zone boundary (and still in one dimension)

It is not too hard to extend this calculation to the case where \mathbf{k} is not quite on a zone boundary. For simplicity though we will stick to the one dimensional situation⁶. We need only solve the secular equation 14.8 for more general k . To do this, we expand around the zone boundaries.

Let us consider the states at the zone boundary $k = \pm n\pi/a$ which are separated by the reciprocal lattice vectors $G = \pm 2n\pi/a$. As noted above, the gap that opens up precisely at the zone boundary will be $\pm |V_G|$. Now let us consider a plane wave near this zone boundary $k = n\pi/a + \delta$ with δ being very small (and n an integer). This wavevector can scatter into $k' = -n\pi/a + \delta$ due to the periodic potential. We then have

$$\begin{aligned}\epsilon_0(n\pi/a + \delta) &= \frac{\hbar^2}{2m} [(n\pi/a)^2 + 2n\pi\delta/a + \delta^2] \\ \epsilon_0(-n\pi/a + \delta) &= \frac{\hbar^2}{2m} [(n\pi/a)^2 - 2n\pi\delta/a + \delta^2]\end{aligned}$$

The secular equation (Eq. 14.8)) is then

$$\left(\frac{\hbar^2}{2m} [(n\pi/a)^2 + \delta^2] - E + \frac{\hbar^2}{2m} 2n\pi\delta/a \right) \left(\frac{\hbar^2}{2m} [(n\pi/a)^2 + \delta^2] - E - \frac{\hbar^2}{2m} 2n\pi\delta/a \right) - |V_G|^2 = 0$$

which simplifies to

$$\left(\frac{\hbar^2}{2m} [(n\pi/a)^2 + \delta^2] - E \right)^2 = \left(\frac{\hbar^2}{2m} 2n\pi\delta/a \right)^2 + |V_G|^2$$

or

$$E_{\pm} = \frac{\hbar^2}{2m} [(n\pi/a)^2 + \delta^2] \pm \sqrt{\left(\frac{\hbar^2}{2m} 2n\pi\delta/a \right)^2 + |V_G|^2} \quad (14.10)$$

⁶If you are very brave and good with geometry, you can try working out the three dimensional case.

Expanding the square root for small δ we obtain⁷

$$E_{\pm} = \frac{\hbar^2(n\pi/a)^2}{2m} \pm |V_G| + \frac{\hbar^2\delta^2}{2m} \left[1 \pm \frac{\hbar^2(n\pi/a)^2}{m} \frac{1}{|V_G|} \right] \quad (14.11)$$

Note that for small perturbation (which is what we are concerned with), the second term in the square brackets is larger than unity so that for one of the two solutions, the square bracket is negative.

Thus we see that near the band gap at the Brillouin zone boundary, the dispersion is quadratic (in δ) as shown in Fig. 14.3. In Fig. 14.4, we see (using the *repeated zone scheme*) that small gaps open at the Brillouin zone boundaries in what is otherwise a parabolic spectrum. (This plotting scheme is equivalent to the reduced zone scheme if restricted to a single zone).

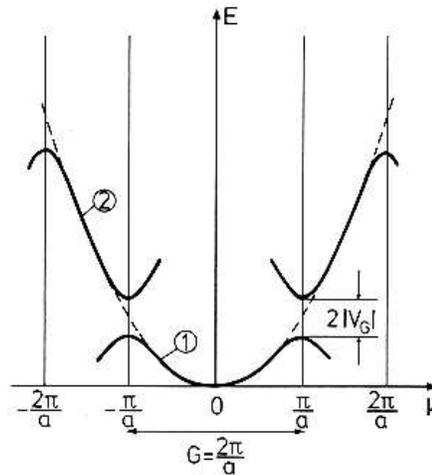


Figure 14.3: Dispersion of a Nearly Free Electron Model. In the nearly free electron model, gaps open up at the Brillouin zone boundaries in an otherwise parabolic spectrum. Compare this to what we found for the tight binding model in Fig 10.5.

The general structure we find is thus very much like what we expected from the tight binding model we considered previously in chapter 10 above. As in the tight binding picture there are *energy bands* where there are energy eigenstates, and there are gaps between bands, where there are no energy eigenstates. As in the tight binding model, the spectrum is periodic in the Brillouin zone (See Fig 14.4).

In section 10.2 above we introduced the idea of the effective mass — if a dispersion is parabolic, we can describe the curvature at the bottom of the band in terms of an effective mass. In this model at every Brillouin zone boundary the dispersion is parabolic (indeed, if there is a

⁷The conditions of validity for this expansion is that the first term under the square root is much smaller than the second, meaning that we must have small enough δ , or we must be very close to the Brillouin zone boundary. But note that as V_G gets smaller and smaller, the expansion is valid only for k closer and closer to the zone boundary.

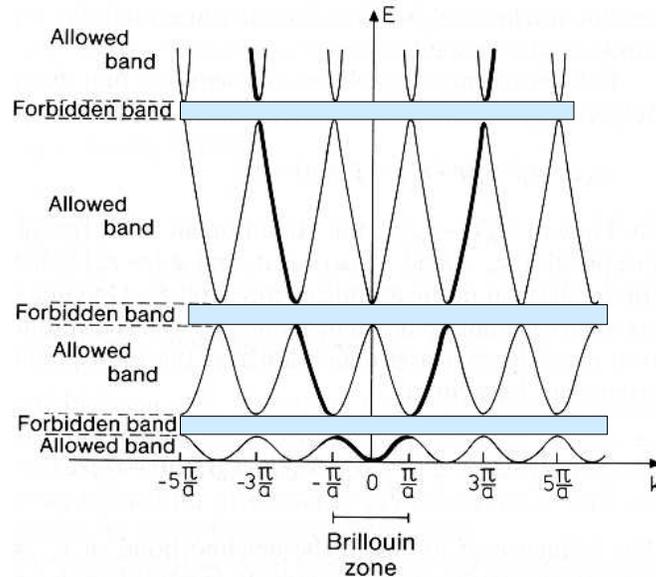


Figure 14.4: Dispersion of a Nearly Free Electron Model. Same as Fig. 14.3 above, but plotted in repeated zone scheme. This is equivalent to the reduced zone scheme but the equivalent zones are repeated. Forbidden bands are marked where there are no eigenstates. The similarity to the free electron parabolic spectrum is emphasized.

gap, hence a local maximum and a local minimum, the dispersion must be parabolic around these extrema). Thus we can write the dispersion Eq. 14.11 as

$$E_+(G + \delta) = C_+ + \frac{\hbar^2 \delta^2}{2m_+^*}$$

$$E_-(G + \delta) = C_- - \frac{\hbar^2 \delta^2}{2m_-^*}$$

where C_+ and C_- are constants, and the effective masses are given here by⁸

$$m_{\pm}^* = \frac{m}{\left| 1 \pm \frac{\hbar^2 (n\pi/a)^2}{m} \frac{1}{|V_G|} \right|}$$

We will define effective mass more precisely, and explain its physics in detail in chapter 16 below. For now we just think of this as a convenient way to describe the parabolic dispersion near the Brillouin zone boundary.

Nearly free electrons in two (and higher) dimensions

The principles of the nearly free electron model are quite similar in two and three dimensions. In short, near the Brillouin zone boundary, a gap opens up due to scattering by a reciprocal lattice vector. States of energy slightly higher than the zone boundary intersection point are pushed up

⁸Note that since V_G is assumed small, $1 - \frac{\hbar^2 (n\pi/a)^2}{m} \frac{1}{|V_G|}$ is negative.

in energy, whereas states of energy slightly lower than the zone boundary intersection point are pushed down in energy. We will return to the detailed geometry of this situation in section 15.2.

There is one more key difference between one dimension and higher dimensions. In one dimension, we found that if \mathbf{k} is on a zone boundary, then there will be exactly one other \mathbf{k}' such that $\mathbf{k} - \mathbf{k}' = \mathbf{G}$ is a reciprocal lattice vector and such that $\epsilon(\mathbf{k}') = \epsilon(\mathbf{k})$. (I.e., Eqs. 14.3 and 14.4 are satisfied). As described above, these two plane wave states mix with each other (See Eq. 14.6) and open up a gap. However, in higher dimensions it may occur that given \mathbf{k} there may be *several* different \mathbf{k}' which will satisfy these equations — i.e., many \mathbf{k}' which differ from \mathbf{k} by a reciprocal lattice vector and which all have the same unperturbed energy. In this case, we need to mix together all of the possible plane waves in order to discover the true eigenstates. One example of when this occurs is the two dimensional square lattice, where the four points $(\pm\pi/a, \pm\pi/a)$ all have the same unperturbed energy and are all separated from each other by reciprocal lattice vectors.

14.2 Bloch's Theorem

In the above, “nearly free electron” approach, we started from the perspective of plane waves that are weakly perturbed by a periodic potential. But in real materials, the scattering from atoms can be very strong so that perturbation theory may not be valid (or may not converge until very high order). How do we know that we can still describe electrons with anything remotely similar to plane waves?

In fact, by this time, after our previous experience with waves, we should know the answer in advance: the plane wave momentum is not a conserved quantity, but the crystal momentum is. No matter how strong the periodic potential, so long as it is periodic, crystal momentum is conserved. This important fact was first discovered by Felix Bloch⁹ in 1928, very shortly after the discovery of the Schrodinger equation, in what has become known as Bloch's theorem¹⁰

Bloch's Theorem: An electron in a periodic potential has eigenstates of the form

$$\Psi_{\mathbf{k}}^{\alpha}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}^{\alpha}(\mathbf{r})$$

where $u_{\mathbf{k}}^{\alpha}$ is periodic in the unit cell and \mathbf{k} (the crystal momentum) can be chosen within the first Brillouin zone.

In reduced zone scheme there may be many states at each \mathbf{k} and these are indexed by α . The periodic function u is usually known as a *Bloch function*, and Ψ is sometimes known as a *modified plane-wave*. Because u is periodic, it can be rewritten as a sum over reciprocal lattice vectors

$$u_{\mathbf{k}}^{\alpha}(\mathbf{r}) = \sum_{\mathbf{G}} \tilde{u}_{\mathbf{G},\mathbf{k}}^{\alpha} e^{i\mathbf{G}\cdot\mathbf{r}}$$

This form guarantees¹¹ that $u_{\mathbf{k}}^{\alpha}(\mathbf{r}) = u_{\mathbf{k}}^{\alpha}(\mathbf{r} + \mathbf{R})$ for any lattice vector \mathbf{R} . Therefore the full

⁹Felix Bloch later won a Nobel prize for inventing Nuclear Magnetic Resonance. NMR was then renamed MRI (Magnetic Resonance Imaging) when people decided the word “Nuclear” sounds too much like it must be related to some sort of bomb.

¹⁰Bloch's theorem was actually discovered by a Mathematician Gaston Floquet in 1883, and rediscovered later by Bloch in the context of solids. This is an example of what is known as Stigler's Law of Eponymy: “Most things are not named after the person who first discovers them”. In fact, Stigler's law was discovered by Merton.

¹¹In fact, the function u is periodic in the unit cell if and only if it can be written as a sum over reciprocal lattice

wavefunction is expressed as

$$\Psi_{\mathbf{k}}^{\alpha}(\mathbf{r}) = \sum_{\mathbf{G}} \tilde{u}_{\mathbf{G},\mathbf{k}}^{\alpha} e^{i(\mathbf{G}+\mathbf{k})\cdot\mathbf{r}} \quad (14.12)$$

Thus an equivalent statement of Bloch's theorem is that we can write each eigenstate as being made up of a sum of plane wave states \mathbf{k} which differ by reciprocal lattice vectors \mathbf{G} .

Given this equivalent statement of Bloch's theorem, we now understand that the reason for Bloch's theorem is that the scattering matrix elements $\langle \mathbf{k}' | V | \mathbf{k} \rangle$ are zero unless \mathbf{k}' and \mathbf{k} differ by a reciprocal lattice vector. As a result, the Schrodinger equation is "block diagonal"¹² in the space of \mathbf{k} and in any given wavefunction only plane waves \mathbf{k} that differ by some \mathbf{G} can be mixed together. One way to see this more clearly is to take the Schrodinger equation

$$\left[\frac{\mathbf{p}^2}{2m} + V(\mathbf{r}) \right] \Psi(\mathbf{r}) = E\Psi(\mathbf{r})$$

and Fourier transform it to obtain

$$\sum_{\mathbf{G}} V_{\mathbf{G}} \Psi_{\mathbf{k}-\mathbf{G}} = \left[E - \frac{\hbar^2 |\mathbf{k}|^2}{2m} \right] \Psi_{\mathbf{k}}$$

where we have used the fact that $V_{\mathbf{k}-\mathbf{k}'}$ is only nonzero if $\mathbf{k} - \mathbf{k}' = \mathbf{G}$. It is then clear that for each \mathbf{k} we have a Schrodinger equation for the set of $\Psi_{\mathbf{k}-\mathbf{G}}$'s and we must obtain solutions of the form of Eq. 14.12.

Although by this time it may not be surprising that electrons in a periodic potential have eigenstates labeled by crystal momenta, we should not overlook how important Bloch's theorem is. This theorem tells us that even though the potential that the electron feels from each atom is extremely strong, the electrons still behave *almost* as if they do not see the atoms at all! They still *almost* form plane wave eigenstates, with the only modification being the periodic Bloch function u and the fact that momentum is now crystal momentum.

A quote from Felix Bloch:

When I started to think about it, I felt that the main problem was to explain how the electrons could sneak by all the ions in a metal. By straight Fourier analysis I found to my delight that the wave differed from the plane wave of free electrons only by a periodic modulation.

14.3 Summary of Electrons in a Periodic Potential

- When electrons are exposed to a periodic potential, gaps arise in their dispersion relation at the Brillouin zone boundary. (The dispersion is quadratic approaching a zone boundary).
- Thus the electronic spectrum breaks into bands, with forbidden energy gaps between the bands. In the nearly free electron model, the gaps are proportional to the periodic potential $|V_{\mathbf{G}}|$.
- Bloch's theorem guarantees us that all eigenstates are some periodic function times a plane wave. In repeated zone scheme the wavevector (the *crystal momentum*) can always be taken in the first Brillouin zone.

vectors in this way.

¹²No pun intended.

References

- Goodstein, Section 3.6a.
- Burns, section 10.1–10.6
- Kittel, chapter 7 (Skip Kronig-Penny model)
- Hook and Hall, section 4.1
- Ashcroft and Mermin, chapters 8–9 (not my favorite)
- Ibach and Luth, sections 7.1–7.2
- Singleton, chapter 2-3

Chapter 15

Insulator, Semiconductor, or Metal

In chapter 10, when we discussed the tight-binding model in one dimension, we introduced some of the basic ideas of band structure. In chapter 14 we found that an electron in a periodic potential shows exactly the same type of band-structure as we found for the tight-binding model: In both cases, we found that the spectrum is periodic in momentum (so all momenta can be taken to be in the first Brillouin zone, in reduced zone scheme) and we find that gaps open at Brillouin zone boundaries. These principles, the idea of bands and band structure form the fundamental underpinning of our understanding of electrons in solids. In this chapter (and the next) we explore these ideas in further depth.

15.1 Energy Bands in One Dimension: Mostly Review

As we pointed out in chapter 12 the number of k -states in a single Brillouin zone is equal to the number of unit cells in the entire system. Thus, if each atom has exactly one electron (i.e., is valence 1) there would be exactly enough electrons to fill the band *if there were only one spin state of the electron*. Being that there are two spin states of the electron, when each atom has only one valence electron, then the band is precisely half full. This is shown in the left of Fig. 15.1. Here, there is a Fermi surface where the unfilled states meet the filled states. (In the figure, the Fermi energy is shown as a green dashed line). When a band is partially filled, the electrons can repopulate when a small electric field is applied, allowing current to flow as shown in the right of Fig. 15.1. Thus, the partially filled band is a *metal*.

On the other hand, if there are two electrons per atom, then we have precisely enough electrons to fill one band. One possibility is shown on the left of Fig. 15.2 — the entire lower band is filled and the upper band is empty, and there is a *band gap* between the two bands (note that the chemical potential is between the bands). When this is the situation, the lower (filled) band is known as the *valence band* and the upper (empty) band is known as the *conduction band*. In this situation the minimum energy excitation is created by moving an electron from the valence to the conduction band, which is nonzero energy. Because of this, at zero temperature, a sufficiently small electric perturbation will not create any excitations— the system does not respond at all to electric field. Thus, systems of this type are known as (electrical) *insulators* (or more specifically

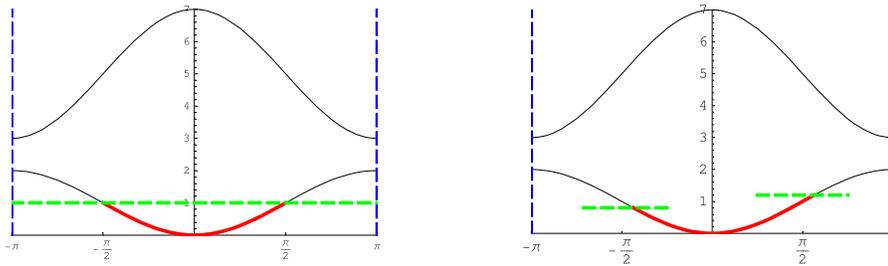


Figure 15.1: Band Diagrams of a One Dimensional Monovalent Chain with Two Orbitals per Unit Cell. Left: A band diagram with two bands shown where each atom has one electron so that the lowest band is exactly half filled, and is therefore a metal. The filled states are colored red, the chemical potential is the green line. Right: When electric field is applied, electrons accelerate, filling some of the k states to the right and emptying k -states to the left (in one dimension this can be thought of as having a different chemical potential on the left versus the right). Since there are an unequal number of left-moving versus right-moving electrons, the situation on the right represents net current flow.

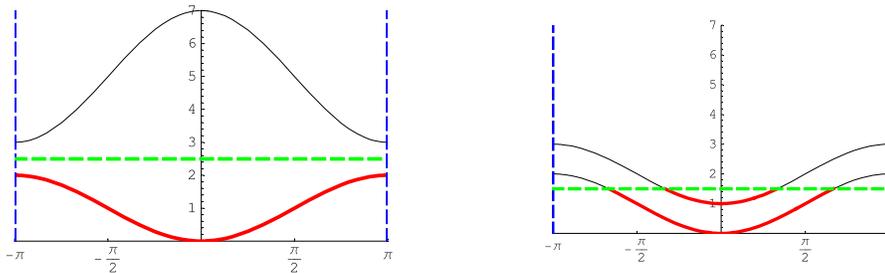


Figure 15.2: Band Diagrams of a One Dimensional Divalent Chain with Two Orbitals per Unit Cell. When there are two electrons per atom, then there are exactly enough electrons to fill the lowest band. In both pictures the chemical potential is drawn in green. Left: one possibility is that the lowest band (the valence band) is completely filled and there is a gap to the next band (the conduction band) in which case we get an insulator. This is a *direct* band gap as the valence band maximum and the conduction band minimum are both at the same crystal momentum (the zone boundary). Right: Another possibility is that the band energies overlap, in which case there are two bands, each of which is partially filled, giving a metal. If the bands were separated by more (imagine just increasing the vertical spacing between bands) we would have an insulator again, this time with an *indirect* band gap, since the valence band maximum is at the zone boundary while the conduction band minimum is at the zone center.

band insulators). If the band gap is below about 4 eV, then these type of insulators are called *semiconductors* since at finite temperature electrons can be thermally excited into the conduction band, and these electrons then can move around freely, carrying some amount of current.

One might want to be aware that in the language of chemists, a band insulator is a situation where all of the electrons are tied up in bonds. For example, in diamond, carbon has valence four — meaning there are four electrons per atom in the outer-most shell. In the diamond lattice, each carbon atom is covalently bonded to each of its four nearest neighbors — and each covalent bond requires two electrons. One electron is donated to each bond from each of the two atoms on either end of the bond — this completely accounts for all of the four electrons in each atom. Thus all of the electrons are tied up in bonds. This turns out to be equivalent to the statement that certain bonding bands are completely filled, and there is no mobility of electrons in any partially filled bands (See the left of Fig. 16.3).

When there are two electrons per atom, one frequently obtains a band insulator as shown in the left of Fig. 15.2. However another possibility is that the band energies overlap, as shown in the right of Fig. 15.2. In this case, although one has precisely the right number of electrons to fill a single band, instead one has two partially filled bands. As in Fig. 15.1 there are low energy excitations available, and the system is metallic.

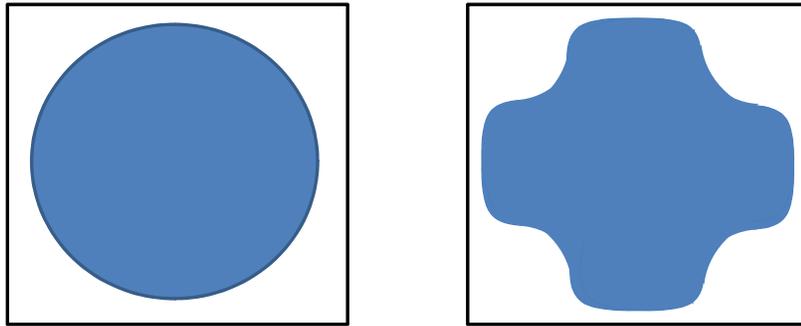


Figure 15.3: Fermi Sea of a Square Lattice of Monovalent Atoms in Two Dimensions. Left: In the absence of a periodic potential, the Fermi sea forms a circle whose area is precisely half that of the Brillouin zone (the black square). Right: when a periodic potential is added, states closer to the zone boundary are pushed down in energy deforming the Fermi sea. Note that the area of the Fermi sea remains fixed.

15.2 Energy Bands in Two (or More) Dimensions

It is useful to try to understand how the nearly-free electron model results in band structure in two dimensions. Let us consider a square lattice of monovalent atoms. The Brillouin zone is correspondingly square, and since there is one electron per atom, there should be enough electrons to half fill a single Brillouin zone. In absence of a periodic potential, the Fermi sea forms a circle as shown in the left of Fig. 15.3. The area of this circle is precisely half the area of the zone. Now when a periodic potential is added, gaps open up at the zone boundaries. This means that states close to the zone boundary get moved down in energy — and the closer they are to the boundary, the more they get moved down. As a result, states close to the boundary get filled up preferentially at the expense of states further from the boundary. This deforms the Fermi surface¹ roughly as shown in the right of Fig. 15.3. In either case, there are low energy excitations possible and therefore the system is a metal.

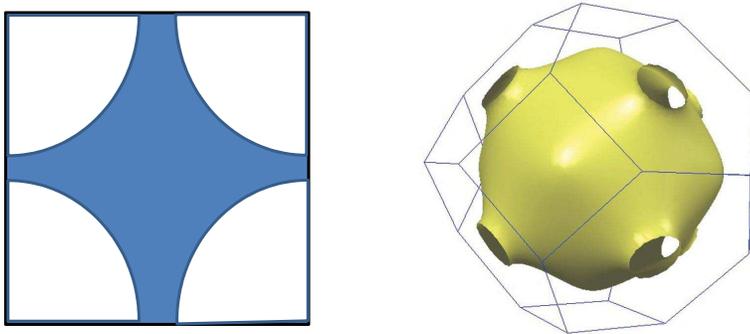


Figure 15.4: Fermi Surfaces that Touch Brillouin Zone Boundaries. Left: Fermi Sea of a square lattice of monovalent atoms in two dimensions with strong periodic potential. The Fermi surface touches the Brillouin zone boundary. Right: The Fermi surface of copper, which is monovalent (the lattice structure is fcc, which determines the shape of the Brillouin zone, see Fig. 12.6).

If the periodic potential is strong enough the Fermi surface may even touch² the Brillouin zone boundary as shown in the left of Fig. 15.4. This is not uncommon in real materials. On the right of Fig. 15.4 the Fermi surface of copper is shown, which similarly touches the zone boundary.

¹Recall that the Fermi surface is the locus of points at the Fermi energy (so all states at the Fermi surface have the same energy), separating the filled from unfilled states. Keep in mind that the area inside the Fermi surface is fixed by the total number of electrons in the system.

²Note that whenever a Fermi surface touches the Brillouin zone boundary, it must do so perpendicularly. This is due to the fact that the group velocity is zero at the zone boundary — i.e., the energy is quadratic as one approaches normal to the zone boundary. Since the energy is essentially not changing in the direction perpendicular to the zone boundary, the Fermi surface must intersect the zone boundary normally.

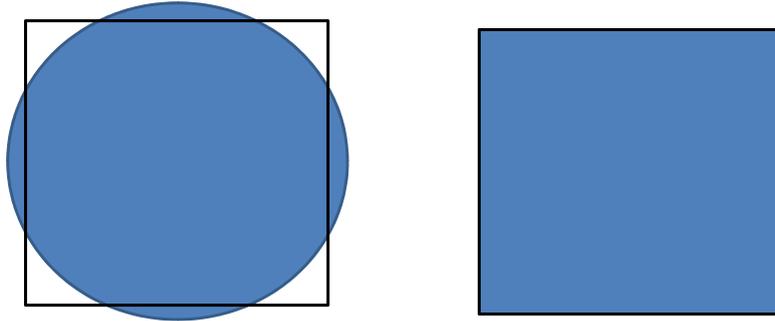


Figure 15.5: Fermi Sea of a Square Lattice of Divalent Atoms in Two Dimensions. Left: In the absence of a periodic potential, the Fermi sea forms a circle whose area is precisely that of the Brillouin zone (the black square). Right: when a *sufficiently strong* periodic potential is added, states inside the zone boundary are pushed down in energy so that all of these states are filled and no states outside of the first Brillouin zone are filled. Since there is a gap at the zone boundary, this situation is an insulator. (Note that the area of the Fermi sea remains fixed).

Let us now consider the case of a two-dimensional square lattice of divalent atoms. In this case the number of electrons is precisely enough to fill a single zone. In the absence of a periodic potential, the Fermi surface is still circular, although it now crosses into the second Brillouin zone, as shown in the left of Fig. 15.5. Again, when a periodic potential is added a gap opens at the zone boundary — this gap opening pushes down the energy of all states within the first zone and pushes up energy of all states in the second zone. If the periodic potential is *sufficiently strong*³, then the states in the first zone are all lower in energy than states in the second zone. As a result, the Fermi sea will look like the right of Fig. 15.5. I.e, the entire lower band is filled, and the upper band is empty. Since there is a gap at the zone boundary, there are no low energy excitations possible, and this system is an insulator.

It is worth considering what happens for intermediate strength of the periodic potential. Again, states outside of the first Brillouin zone are raised in energy and states inside the first Brillouin zone are lowered in energy. Therefore fewer states will be occupied in the second zone and more states occupied in the first zone. However, for intermediate strength of potential, there will remain some states occupied in the second zone and some states empty within the first zone. This is precisely analogous to what happens in the right half of Fig. 15.2. Analogously, there will

³We can estimate how strong the potential needs to be. We need to have the highest energy state in the first Brillouin zone be lower energy than the lowest energy state in the second zone. The highest energy state in the first zone, in the absence of periodic potential, is in the zone corner and therefore has energy $\epsilon_{corner} = 2(\pi/2)^2/(2m)$. The lowest energy state in the second zone is at the middle of the zone boundary edge and in the absence of periodic potential has energy $\epsilon_{edge} = (\pi/2)^2/(2m)$. Thus we need to open up a gap at the zone boundary which is sufficiently large that the edge becomes higher in energy than the corner. This requires roughly that $2|V_G| = \epsilon_{corner} - \epsilon_{edge}$.

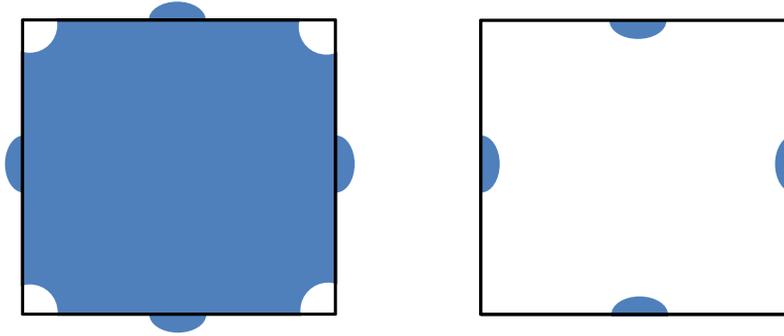


Figure 15.6: Fermi Sea of a Square Lattice of Divalent Atoms in Two Dimensions. Left: For intermediately strong periodic potential, there are still some states filled in the second zone, and some states empty in the first zone, thus the system is still a metal. Right: The states in the second zone can be moved into the first zone by translation by a reciprocal lattice vector. This is the reduced zone scheme representation of the occupancy of the second Brillouin zone.

still be some low energy excitations available, and the system remains a metal.

We emphasize that in the case where there are many atoms per unit cell, we should count the *total valence* of all of the atoms in the unit cell put together to determine if it is possible to obtain a filled-band insulator. If the total valence in of all the atoms in the unit cell is even, then for strong enough periodic potential, it is possible that some set of low energy bands will be completely filled, there will be a gap, and the remaining bands will be empty – i.e., it will be a band insulator.

15.3 Tight Binding

So far in this chapter we have described band structure in terms of the nearly free electron model. Similar results can be obtained starting from the opposite limit — the tight binding model introduced in chapter 10. In this model we imagine some number of orbitals on each atom (or in each unit cell) and allow them to only weakly hop between orbitals. This spreads the eigen-energies of the atomic orbitals out into bands.

Writing down a two (or three) dimensional generalization of the tight binding Hamiltonian Eq. 10.4 is quite straightforward and is a good exercise to try. One only needs to allow each orbital to hop to neighbors in all available directions. The eigenvalue problem can then always be solved with a plane wave ansatz analogous to Eq. 10.5. The solution (again a good exercise to try!) of a tight binding model of atoms, each having a single atomic orbital, on a square lattice is given by

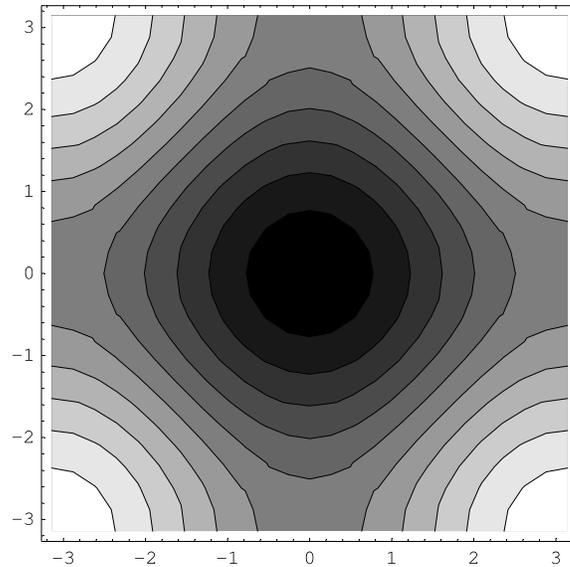


Figure 15.7: Equi-Energy Contours for the Dispersion of a Tight Binding Model on a Square Lattice. This is a contour plot of Eq. 15.1. The first Brillouin Zone is shown. Note that the contours intersect the Brillouin zone boundary normally.

(Compare Eq. 10.6)

$$E(k) = \epsilon_0 - 2t \cos(k_x a) - 2t \cos(k_y a) \quad (15.1)$$

Equi-energy contours for this expression are shown in Fig. 15.7. Note the similarity in the dispersion to our qualitative expectations shown in Fig. 15.3 (right) and Fig. 15.4 left and Fig. 15.6, which were based on a nearly free electron picture.

In the above described tight binding picture, there is only a single band. However, one can make the situation more realistic by starting with several atomic orbitals per unit cell, to obtain several bands (another good exercise to try!). As mentioned above in section 5.3.2 and chapter 10, as more and more orbitals are added to a tight binding (or LCAO) calculation, the results become increasingly accurate.

In the case where a unit cell is divalent, as mentioned above, it is crucial to determine whether bands overlap. (I.e., is it insulating like the left of Fig. 15.2 or metallic type like the right of Fig. 15.2.) This, of course, requires detailed knowledge of the band structure. In the tight binding picture, if the atomic orbitals start sufficiently far apart in energy, then small hopping between atoms cannot spread the bands enough to make them overlap (See Fig. 10.4). In the nearly free electron picture, the gap between bands formed at the Brillouin zone boundary is proportional to $|V_{\mathbf{G}}|$, and it is the limit of strong periodic potential that will guarantee that the bands do not overlap (See Fig. 15.5). Qualitatively these two are the same limit — very far from the idea of a freely propagating wave!

15.4 Failures of the Band-Structure Picture of Metals and Insulators

The picture we have developed is that the band structure, and the filling of bands, determines whether a material is a metal or insulator, (or semiconductor, meaning an insulator with a small band gap). One thing we might conclude at this point is that any system where the unit cell has a single valence electron (so the first Brillouin zone is half-full) must be a metal. However, it turns out that this is not always true! The problem is that we have left out a very important effect — Coulomb interaction between electrons. We have so far completely ignored the Coulomb repulsion between electrons. Is this neglect justified at all? If we try to estimate how strong the Coulomb interaction is between electrons, (roughly $e^2/(4\pi\epsilon_0 r)$ where r is the typical distance between two electrons — i.e., the lattice constant a) we find numbers on the order of several eV. This number can be larger, or even far larger, than the Fermi energy (which is already a very large number, on the order of 10,000 K). Given this, it is hard to explain why it is at all justified to have thrown out such an important contribution. In fact, one might expect that neglecting this term would give complete nonsense! Fortunately, it turns out that in many cases it *is* OK to assume noninteracting electrons. The reason this works is actually quite subtle and was not understood until the 1950s due to the work of Lev Landau (See footnote 12 in chapter 4 about Landau). This (rather deep) explanation, however, is beyond the scope of this course so we will not discuss it. Nonetheless, with this in mind it is perhaps not too surprising that there are cases where the noninteracting electron picture, and hence our view of band structure, fails.

Magnets

A case where the band picture of electrons fails is when the system is ferromagnetic⁴. We will discuss ferromagnetism in detail in chapters 19–22 below, but in short this is where, due to interaction effects, the electron spins spontaneously align. From a kinetic energy point of view this seems unfavorable, since filling the lower energy states with two spins can lower the Fermi energy. However, it turns out that aligning all of the spins can lower the Coulomb energy between the electrons, and thus our rules of non-interacting electron band theory no longer hold.

Mott Insulators

Another case where interaction physics is important is the so-called Mott insulator⁵. Consider a monovalent material. From band theory one might expect a half-filled lowest band, therefore a metal. But if one considers the limit where the electron-electron interaction is extremely strong, this is not what you get. Instead, since the electron-electron interaction is very strong, there is a huge penalty for two electrons to be on the same atom (even with opposite spins). As a result, the ground state is just one electron sitting on each atom. Since each atom has exactly one electron, no electron can move from its atom — since that would result in a double occupancy of the atom it lands on. As a result, this type of ground state is insulating. In some sense this type of insulator — which can be thought of as more-or-less a traffic jam of electrons — is actually simpler to visualize than a band insulator! We will also discuss Mott insulators further in sections 18.4 and particularly 22.2 below.

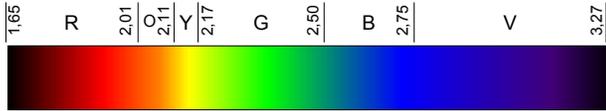
⁴Or antiferromagnetic or ferrimagnetic, for that matter. See chapter 19 below for definitions of these terms.

⁵Named after the English Nobel Laureate, Nevill Mott. Classic examples of Mott insulators include NiO and CoO.

15.5 Band Structure and Optical Properties

To the extent that electronic band structure is a good description of the properties of materials (and usually it is), one can attribute many of the optical properties of materials to this band structure. First one needs to know a few simple facts about light shown here in this table:

Color	$\hbar\omega$
Infrared	< 1.65 eV
Red	~ 1.8 eV
Orange	~ 2.05 eV
Yellow	~ 2.15 eV
Green	~ 2.3 eV
Blue	~ 2.7 eV
Violet	~ 3.1 eV
Ultraviolet	> 3.2 eV



15.5.1 Optical Properties of Insulators and Semiconductors

With this table in mind we see that if an insulator (or wide-bandgap semiconductor) has a bandgap of greater than 3.2 eV, then it appears transparent. The reason for this is that a single photon of visible light cannot excite an electron in the valence band into the conduction band. Since the valence band is completely filled, the minimum energy excitation is of the band gap energy — so the photon creates no excitations at all. As a result, the visible optical photons do not scatter from this material at all and they simply pass right through the material⁶. Materials such as quartz, diamond, aluminum-oxide, and so forth are insulators of this type.

Semiconductors with somewhat smaller band gaps will absorb photons with energies above the band gap (exciting electrons from the valence to the conduction band), but will be transparent to photons below this band gap. For example, cadmium-sulfide (CdS) is a semiconductor with band gap of roughly 2.6 eV, so that violet and blue light are absorbed but red and green light are transmitted. As a result this material looks reddish. (See Fig. 15.8).

15.5.2 Direct and Indirect Transitions

While the band gap determines the minimum energy excitation that can be made in an insulator (or semiconductor), this is not the complete story in determining whether or not a photon can be absorbed by a material. It turns out to matter quite a bit at which values of \mathbf{k} the maximum of the valence band and the minimum of the conduction band lies. If the value of \mathbf{k} for the valence band maximum is the same as the value of \mathbf{k} for the conduction band minimum, then we say that it is a *direct* band gap. If the values of \mathbf{k} differ, then we say that it is an *indirect* band gap. For example, the system shown on the left of Fig. 15.2 is a direct band gap, where both the valence band maximum and the conduction band minimum are at the zone boundary. In comparison, if the band shapes were as in the right of Fig. 15.2, but the band gap were large enough such that it would be an insulator (just imagine the bands separated by more), this would be an indirect band gap since the valence band maximum is at the zone boundary, but the conduction band minimum is at $\mathbf{k} = \mathbf{0}$.

⁶Very weak scattering processes can occur where, say, two photons together can excite an electron, or a photon excites a phonon



Figure 15.8: Orange crystals of CdS. This particular crystal is the naturally occurring mineral called “Greenockite” which is CdS with trace amounts of impurity which can change its color somewhat.

One can also have both indirect and direct band gaps in the same material, as shown in Fig. 15.9. In this figure, the minimum energy excitation is the *indirect transition* — meaning an excitation of an electron across an indirect band gap, or equivalently a transition of nonzero crystal momentum⁷ where the electron is excited from the top of the valence band to the bottom of the lower conduction band at a very different \mathbf{k} . While this may be the lowest energy excitation that can occur, it is very hard for this type of excitation to result from exposure of the system to light — the reason for this is energy-momentum conservation. If a photon is absorbed, the system absorbs both the energy and the momentum of the photon. But given an energy E in the eV range, the momentum of the photon $|k| = E/c$ is extremely small, because c is so large. Thus the system cannot conserve momentum while exciting an electron across an indirect band gap. Nonetheless, typically if a system like this is exposed to photons with energy greater than the indirect band gap a small number of electrons will manage to get excited — usually by some complicated process including absorption of a photon exciting an electron with simultaneous emission of a phonon⁸ to arrange the conservation of energy and momentum. In comparison, if a system has a direct band gap, and is exposed to photons of energy matching this direct band gap, then it strongly absorbs these photons while exciting electrons from the valence band to the conduction band.

15.5.3 Optical Properties of Metals

The optical properties of metals, however, are a bit more complicated. Since these materials are very conductive, photons (which are electromagnetic) excite the electrons⁹, which then re-emit light. This re-emission (or reflection) of light is why metals look shiny. Noble metals (gold, silver,

⁷By “nonzero” we mean, substantially nonzero — like a fraction of the Brillouin zone.

⁸Another way to satisfy the conservation of momentum is via a “disorder assisted” process. You recall that the reason we conserve crystal momentum is because the system is perfectly periodic. If the system has some disorder, and is therefore not perfectly periodic, then crystal momentum is not perfectly conserved. Thus the greater the disorder level, the less crystal momentum needs to be conserved and the easier it is to make a transition across an indirect band gap.

⁹Note the contrast with insulators — when an electron is excited above the band gap, since the conductivity is somewhat low, the electron does not re-emit quickly, and the material mostly just absorbs the given wavelength.

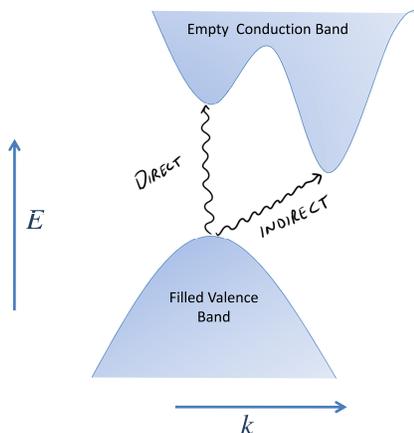


Figure 15.9: Direct and Indirect transitions. While the indirect transition is lower energy, it is hard for a photon to excite an electron across an indirect band gap because photons carry very little momentum (since the speed of light, c , is large).

platinum) look particularly shiny because their surfaces do not form insulating oxides when exposed to air, which many metals (such as sodium) do within seconds.

Even amongst metals (ignoring possible oxide surfaces), colors vary. For example, Silver looks brighter than gold and copper, which look yellow or orange-ish. This again is a result of the band structure of these materials. Both of these materials have valence one meaning that a band should be half-filled. However, the total energy width of the conduction band is greater for silver than it is for gold or copper (In tight-binding language t is larger for silver, see chapter 10). This means that higher energy electronic transitions within the band are much more possible for silver than they are for gold and copper. For copper and gold, photons with blue and violet colors are not well absorbed and re-emitted, leaving these material looking a bit more yellow and orange. For silver on the other hand, all visible colors are re-emitted well, resulting in a more perfect (or “white”) mirror. While this discussion of the optical properties of metals is highly oversimplified¹⁰, it captures the correct essence — that the details of the band structure determine which color photons are easily absorbed and/or reflected, and this in turn determines the apparent color of the material.

15.5.4 Optical Effects of Impurities

It turns out that small levels of impurities put into periodic crystals (particularly into semiconductors and insulators) can have dramatic effects on many of their optical (as well as electrical!) properties. For example, one nitrogen impurity per million carbon atoms in a diamond crystal gives the crystal a yellow-ish color. One boron atom per million carbon atoms give the diamond a blue-ish color¹¹. We will discuss the physics that causes this in section 16.2.1 below.

¹⁰Really there are many bands overlapping in these materials and the full story addresses inter and intra-band transitions.

¹¹Natural blue diamonds are extremely highly prized and are very expensive. Possibly the world’s most famous diamond, the Hope Diamond, is of this type (it is also supposed to be cursed, but that is another story). With modern crystal growth techniques, in fact it is possible to produce man-made diamonds of “quality” better than those that are mined. Impurities can be placed in as desired to give the diamond any color you like. Due to the

15.6 Summary of Insulators, Semiconductors, and Metals

- A material is a metal if it has low energy excitations. This happens when at least one band is partially full. (Band) Insulators and semiconductors have only filled bands and empty bands and have a gap for excitations.
- A semiconductor is a (band) insulator with a small band gap.
- The valence of a material determines the number of carriers being put into the band — and hence can determine if one has a metal or insulator/semiconductor. However, if bands overlap (and frequently they do) one might not be able to fill the bands to a point where there is a gap.
- The gap between bands is determined by the strength of the periodic potential. If the periodic potential is strong enough (the atomic limit in tight binding language), bands will not overlap.
- The band picture of materials fails to account for electron-electron interaction. It cannot describe (at least without modification) interaction driven physics such as magnetism and Mott insulators.
- Optical properties of solids depend crucially on the possible energies of electronic transitions. Photons easily create transitions with low momentum, but cannot create transitions with larger momentum easily. Optical excitations over an indirect (finite momentum) gap are therefore weak.

References

- Goodstein section 3.6c
- Kittel, chapter 7; first section of chapter 8; first section of chapter 9
- Burns, section 10.7, 10.10
- Hook and Hall, section 4.2,4.3, section 5.4
- Rosenberg section 8.9–8.19

powerful lobby of the diamond industry, all synthetic diamonds are labeled as such — so although you might feel cheap wearing a synthetic, in fact, you probably own a better product than those that have come out of the earth! (Also you can rest with a clean conscience that the production of this diamond did not finance any wars in Africa).

Chapter 16

Semiconductor Physics

16.1 Electrons and Holes

Suppose we start with an insulator or semiconductor and we excite one electron from the valence band to the conduction band, as shown in the left of Fig. 16.1. This excitation may be due to absorbing a photon, or it might be a thermal excitation. (For simplicity in the figure we have shown a direct band gap. For generality we have not assumed that the curvature of the two bands are the same). When the electron has been moved up to the conduction band, there is an absence of an electron in the valence band known as a *hole*. Since a completely filled band is inert, it is very convenient to only keep track of the few holes in the valence band (assuming there are only a few) and to treat these holes as individual elementary particles. The electron can fall back into the empty state that is the hole, emitting energy (a photon, say) and “annihilating” both the electron from the conduction band and the hole from the valence band¹. Note that while the electrical charge of an electron is negative the electrical charge of a hole (the absence of an electron) is positive — equal and opposite to that of the electron.²

Effective Mass of Electrons

As mentioned in sections 10.2 and 14.1.1, it is useful to describe the curvature at the bottom of a band in terms of an effective mass. Let us assume that near the bottom of the conduction band

¹This is equivalent to pair annihilation of an electron with a positron. In fact, the analogy between electron-hole and electron-positron is fairly precise. As soon as Dirac constructed his equation (in 1928) describing the relativistic motion of electrons, and predicting positrons, it was understood that the positron could be thought of as an absence of an electron in an filled sea of states. The filled sea of electron states with a gap to exciting electron-positron pairs is the inert vacuum, which is analogous to an inert filled valence band.

²If this does not make intuitive sense consider the process of creating an electron-hole pair as described in Fig. 16.1. Initially (without the excited electron-hole pair) the system is charge neutral. We excite the system with a photon to create the pair, and we have not moved any additional net charge into the system. Thus if the electron is negative, the hole must be positive to preserve overall charge neutrality.

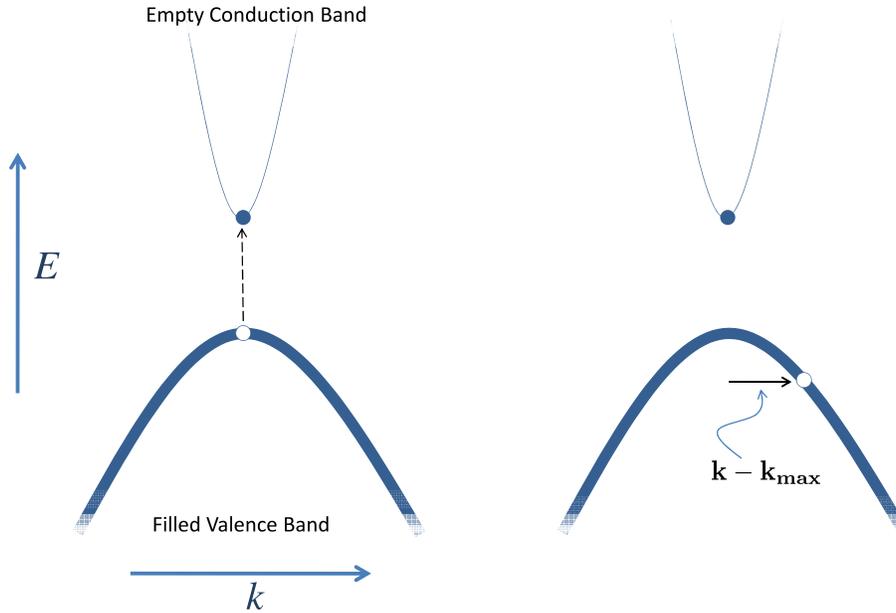


Figure 16.1: Electrons and Holes in a Semiconductors. Left: A single hole in the valence band and a single electron in the conduction band. Right: Moving the hole to a momentum away from the top of the valence band costs *positive* energy — like pushing a balloon under water. As such, the effective mass of the hole is defined to be positive. The energy of the configuration on the right is greater than that on the left by $E = \hbar^2|\mathbf{k} - \mathbf{k}_{\max}|^2/(2m^*)$

(assumed to be at $\mathbf{k} = \mathbf{k}_{\min}$) the energy is given by^{3,4,5}

$$E = E_{min} + \alpha|\mathbf{k} - \mathbf{k}_{\min}|^2 + \dots$$

where the dots mean higher order term in the deviation from \mathbf{k}_{\min} . We then *define* the *effective mass* to be given by

$$\frac{\hbar^2}{m^*} = \frac{\partial^2 E}{\partial k^2} = 2\alpha \quad (16.1)$$

at the bottom of the band (with the derivative being taken in any direction for an isotropic system). Correspondingly, the (group) velocity is given by

$$\mathbf{v} = \nabla_{\mathbf{k}} E / \hbar = \hbar(\mathbf{k} - \mathbf{k}_{\min}) / m^* \quad (16.2)$$

³It is an important principle that near a minimum or a maximum one can always expand and get something quadratic plus higher order corrections.

⁴For simplicity we have assumed the system to be isotropic. In the more general case we would have

$$E = E_{min} + \alpha_x(k_x - k_x^{min})^2 + \alpha_y(k_y - k_y^{min})^2 + \alpha_z(k_z - k_z^{min})^2 + \dots$$

for some orthogonal set of axes (the “principle axes”) x, y, z . In this case we would have an effective mass which can be different in the three different principle directions.

⁵For simplicity we also neglect the spin of the electron here. In general, spin-orbit coupling can make the dispersion depend on the spin state of the electron. Among other things, this can modify the effective electron g -factor.

This definition is chosen to be in analogy with the free electron behavior $E = \frac{\hbar^2 |\mathbf{k}|^2}{2m}$ with corresponding velocity $\mathbf{v} = \nabla_{\mathbf{k}} E / \hbar = \hbar \mathbf{k} / m$.

Effective Mass of Holes

Analogously we can define an effective mass for holes. Here things get a bit more complicated⁶. For the top of the valence band, the energy dispersion for electrons would be

$$E = E_{\max} - \alpha |\mathbf{k} - \mathbf{k}_{\max}|^2 + \dots$$

The modern convention is to *define the effective mass for holes at the top of a valence band to be always positive*⁷

$$\frac{\hbar^2}{m_{\text{hole}}^*} = \left| \frac{\partial^2 E}{\partial k^2} \right| = 2\alpha \quad (16.3)$$

The convention of the effective mass being positive makes sense because the energy to boost the hole from zero velocity ($\mathbf{k} = \mathbf{k}_{\max}$ at the top of the valence band) to finite velocity is positive. This energy is naturally given by

$$E_{\text{hole}} = \frac{\hbar^2 |\mathbf{k} - \mathbf{k}_{\max}|^2}{2m_{\text{hole}}^*}$$

The fact that boosting the hole away from the top of the valence band is positive energy may seem a bit counter-intuitive being that the dispersion of the hole band is an upside-down parabola. However, one should think of this like pushing a balloon under water. The lowest energy configuration is with the *electrons* at the lowest energy possible and the hole at the highest energy possible. So pushing the hole under the electrons costs positive energy. (This is depicted in the right hand side of Fig. 16.1.)

Analogous to the electron, we can write the hole group velocity as the derivative of the hole energy

$$\mathbf{v}_{\text{hole}} = \nabla_{\mathbf{k}} E_{\text{hole}} / \hbar = \hbar (\mathbf{k} - \mathbf{k}_{\max}) / m_{\text{hole}}^* \quad (16.4)$$

Effective Mass and Equations of Motion

We have defined the effective masses above in analogy with that of free electrons, by looking at the curvature of the dispersion. An equivalent definition (equivalent at least at the top or bottom of the band) is to define the effective mass m^* as being the quantity that satisfies Newton's second law, $F = m^* a$ for the particle in question. To demonstrate this, our strategy is to imagine applying a force to an electron in the system and then equate the work done on the electron to its change in energy. Let us start with an electron in momentum state \mathbf{k} . Its group velocity is $\mathbf{v} = \nabla_{\mathbf{k}} E(\mathbf{k}) / \hbar$. If we apply a force⁸, the work done per unit time is

$$dW/dt = \mathbf{F} \cdot \mathbf{v} = \mathbf{F} \cdot \nabla_{\mathbf{k}} E(\mathbf{k}) / \hbar$$

⁶Some people find the concept of effective mass for holes to be a bit difficult to digest. I recommend chapter 12 of Ashcroft and Mermin to explain this in more detail (in particular see page 225 and thereafter).

⁷Be warned: a few books define the mass of holes to be negative. This is a bit annoying but not inconsistent as long as the negative sign shows up somewhere else!

⁸For example, if we apply an electric field \mathbf{E} and it acts on an electron of charge $-e$, the force is $\mathbf{F} = -e\mathbf{E}$.

On the other hand, the change in energy per unit time must also be (by the chain rule)

$$dE/dt = d\mathbf{k}/dt \cdot \nabla_{\mathbf{k}}E(\mathbf{k})$$

Setting these two expressions equal to each other we (unsurprisingly) obtain Newton's equation

$$\mathbf{F} = \frac{1}{\hbar} \frac{d\mathbf{k}}{dt} = \frac{d\mathbf{p}}{dt} \quad (16.5)$$

where we have used $\mathbf{p} = \hbar\mathbf{k}$.

If we now consider electrons near the bottom of a band, we can plug in the expression Eq. 16.2 for the velocity and this becomes

$$\mathbf{F} = m^* \frac{d\mathbf{v}}{dt}$$

exactly as Newton would have expected. In deriving this result recall that we have assumed that we are considering an electron near the bottom of a band so that we can expand the dispersion quadratically (or similarly we assumed that holes are near the top of a band). One might wonder how we should understand electrons when they are neither near the top nor the bottom of a band. More generally Eq. 16.5 always holds, as does the fact that the group velocity is $\mathbf{v} = \nabla_{\mathbf{k}}E/\hbar$. It is then sometimes convenient to define an effective mass for an electron as a function of momentum to be given by⁹

$$\frac{\hbar^2}{m^*(k)} = \frac{\partial^2 E}{\partial k^2}$$

which agrees with our above definition (Eq. 16.1) near the bottom of band. However, near the top of a band it is the *negative* of the corresponding hole mass (note the absolute value in Eq. 16.3). Note also that somewhere in the middle of the band the dispersion must reach an inflection point ($\partial^2 E/\partial k^2 = 0$), whereupon the effective mass actually becomes infinite as it changes sign.

Aside: It is useful to compare the time evolution of electrons and holes near the top of bands. If we think in terms of holes (the natural thing to do near the top of a band) we have $\mathbf{F} = +e\mathbf{E}$ and the holes have a positive mass. However if we think in terms of electrons, we have $\mathbf{F} = -e\mathbf{E}$ but the mass is negative. Either way, the acceleration of the \mathbf{k} -state is the same, whether we are describing the state in terms of an electron in the state or in terms of a hole in the state. This is a rather important fundamental principle — that the time evolution of an eigenstate is independent of whether that eigenstate is filled with an electron or not.

16.1.1 Drude Transport: Redux

Back in section 3 we studied Drude theory — a simple kinetic theory of electron motion. The main failure of Drude theory was that it did not treat the Pauli exclusion principle properly: it neglected the fact that in metals the high density of electrons makes the Fermi energy extremely high. However, in semiconductors or band insulators, when only a few electrons are in the conduction band and/or only a few holes are in the valence band, then we can consider this to be a low density situation, and to a very good approximation, we can ignore Fermi statistics. (For example, if only a single electron is excited into the conduction band, then we can completely ignore the Pauli principle, since it is the only electron around — there is no chance that any state it wants to sit in will already be filled!). As a result, when there is a low density of conduction electrons or valence holes, it turns out that Drude theory works extremely well! We will come back to this issue later in section 16.3 and make this statement much more precise.

⁹For simplicity we write this in its one dimensional form.

At any rate, in the semiclassical picture, we can write a simple Drude transport equation (really Newton's equations!) for electrons in the conduction band

$$m_e^* d\mathbf{v}/dt = -e(\mathbf{E} + \mathbf{v} \times \mathbf{B}) - m_e^* \mathbf{v}/\tau$$

with m_e^* the electron effective mass. Here the first term on the right hand side is the force on the electron, and the second term is a drag force with an appropriate scattering time τ . The scattering time determines the so-called *mobility* μ which measures the ease with which the particle moves¹⁰

$$\mu = |\mathbf{E}|/|\mathbf{v}| = |e\tau/m^*|$$

Similarly we can write equations of motion for holes in the valence band

$$m_h^* d\mathbf{v}/dt = e(\mathbf{E} + \mathbf{v} \times \mathbf{B}) - m_h^* \mathbf{v}/\tau$$

where m_h^* is the hole effective mass. Note again that here the charge on the hole is *positive*. This should make sense — the electric field pulls on the electrons in a direction opposite than it pulls on the absence of an electron!

If we think back all the way to chapter 3 and 4, one of the physical puzzles that we could not understand is why the Hall coefficient sometimes changes sign (See the table in section 3.1.2). In some cases it looked as if the charge carrier had positive charge. Now we understand why this is true. In some materials the main charge carrier is the hole!

16.2 Adding Electrons or Holes With Impurities: Doping

In a pure band insulator or semiconductor, if we excite electrons from the valence to the conduction band (either with photons or thermally) we can be assured that the number of electrons in the conduction band (typically called n) is precisely equal to the number of holes left behind in the valence band (typically called p). However, in an impure semiconductor or band insulator this is not the case.

Consider for example, silicon (Si), which is a semiconductor with a band gap of about 1.1 eV. Without impurities, a semiconductor is known as *intrinsic*¹¹. Now imagine that a phosphorus (P) atom replaces one of the Si atoms in the lattice as shown on the left of Fig. 16.2. This P atom, being directly to the right of Si on the periodic table, can be thought of as nothing more than a Si atom plus an extra proton and an extra electron¹² as shown on the right of Fig. 16.2. Since the valence band is already filled this additional electron must go into the conduction band. The P atom is known as a *donor* (or *electron donor*) in silicon since it donates an electron to the conduction band. It is also sometimes known as an n-dopant, since n is the symbol for the density of electrons in the conduction band.

Analogously, we can consider aluminum, the element directly to the left of Si on the periodic table. In this case, the aluminum dopant provides one fewer electron than Si, so there will be one missing electron from the valence band. In this case Al is known as an electron *acceptor*, or equivalently as a p-dopant, since p is the symbol for the density of holes¹³.

¹⁰Mobility is defined to be positive for both electrons and holes.

¹¹The opposite of intrinsic, the case where impurities donate carriers is sometimes known as *extrinsic*.

¹²There is an extra neutron as well, but it doesn't do much in this context.

¹³Yes, it is annoying that the common dopant phosphorus, has the chemical symbol P, but it is not a *p*-dopant, it is an *n*-dopant.

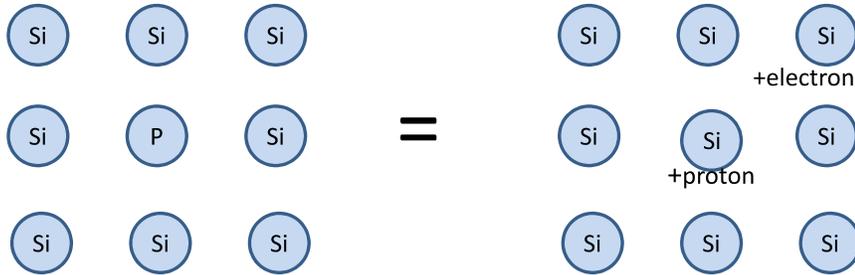


Figure 16.2: Cartoon of Doping a Semiconductor. Doping Si with P adds one free electron

In a more chemistry oriented language, we can depict the donors and acceptors as shown in Fig. 16.3. In the intrinsic case, all of the electrons are tied up in covalent bonds of two electrons. With the n-dopant, there is an extra unbound electron, whereas with the p-dopant there is an extra unbound hole (one electron too few).

16.2.1 Impurity States

Let us consider even more carefully what happens when we add dopants. For definiteness let us consider adding an n-dopant such as P to a semiconductor such as Si. Once we add a single n-dopant to an otherwise intrinsic sample of Si, we get a single electron above the gap in the conduction band. This electron behaves like a free particle with mass m_e^* . However, in addition, we have a single extra positive charge $+e$ at some point in the crystal due to the P nucleus. The free electron is attracted back to this positive charge and forms a bound state that is just like a hydrogen atom. There are two main differences between a real hydrogen atom and this bound state of an electron in the conduction band and the impurity nucleus. First of all, the electron has effective mass m_e^* which can be very different from the real (bare) mass of the electron (and is typically smaller than the bare mass of the electron). Secondly, instead of the two charges attracting each other with a potential $V = \frac{e^2}{4\pi\epsilon_0 r}$ they attract each other with a potential $V = \frac{e^2}{4\pi\epsilon_r\epsilon_0 r}$ where ϵ_r is the relative permittivity (or relative dielectric constant) of the material. With these two small differences, we can calculate the hydrogenic bound states proceeds exactly as we do for genuine hydrogen in our quantum mechanics courses.

We recall the energy eigenstates of the hydrogen atom are given by $E_n^{H-atom} = -Ry/n^2$

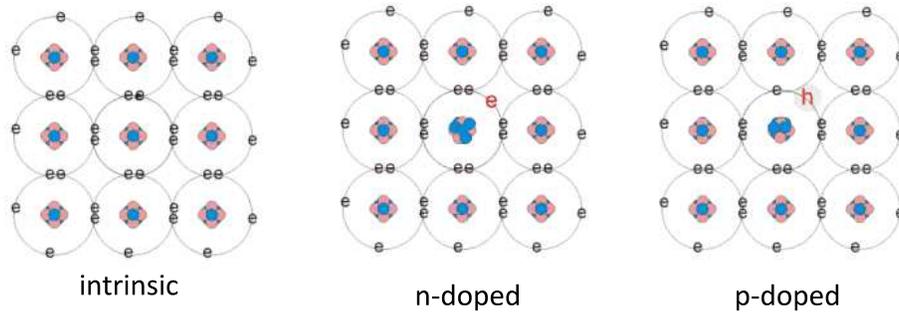


Figure 16.3: Cartoon of Doping a Semiconductor. n and p doping: In the intrinsic case, all of the electrons are tied up in covalent bonds of two electrons. In the n-dopant case, there is an extra unbound electron, whereas with the p-dopant there is an extra hole.

where Ry is the Rydberg constant given by

$$Ry = \frac{me^2}{8\epsilon_0^2\hbar^2} \approx 13.6\text{eV}$$

with m the electron mass. The corresponding radius of this wavefunction is $r_n \approx n^2a_0$ with the Bohr radius given by

$$a_0 = \frac{4\pi\epsilon_0\hbar^2}{me^2} \approx .51 \times 10^{-10}\text{m}$$

The analogous calculation for a hydrogenic impurity state in a semiconductor gives precisely the same expression, only ϵ_0 is replaced by $\epsilon_0\epsilon_r$ and m is replaced by m_e^* . One obtains

$$Ry^{\text{eff}} = Ry \left(\frac{m_e^*}{m} \frac{1}{\epsilon_r^2} \right)$$

and

$$a_0^{\text{eff}} = a_0 \left(\epsilon_r \frac{m}{m_e^*} \right)$$

Because the dielectric constant of semiconductors is typically high (roughly 10 for most common semiconductors) and because the effective mass is frequently low (a third of m or even smaller), the effective Rydberg Ry^{eff} can be tiny compared to the real Rydberg, and the effective

Bohr radius a_0^{eff} can be huge compared to the real Bohr radius¹⁴. For example, in Silicon¹⁵ the effective Rydberg, Ry^{eff} , is much less than .1 eV and a_0^{eff} is above 30 angstroms! Thus this donor impurity forms an energy eigenstate just below the conduction band. At zero temperature this eigenstate will be filled, but it takes only a small temperature to excite some of the bound electrons out of the hydrogenic orbital and into the conduction band.

A depiction of this physics is given in Fig. 16.4 where we have plotted an energy diagram for a semiconductor with donor or acceptor impurities. Here the energies eigenstates are plotted as a function of position. Between the valence and conduction band (which are uniform in position), there are many localized hydrogen-atom-like eigenstates. The energies of these states are in a range of energies but are not all exactly the same since each impurity atom is perturbed by other impurity atoms in its environment. If the density of impurities is high enough, electrons (or holes) can hop from one impurity to the next, forming an *impurity band*.

Note that because the effective Rydberg is very small, the impurity eigenstates are only slightly below the conduction band or above the valence band respectively. With a small temperature, these donors or acceptors can be thermally excited into the band. Thus, except at low enough temperature that the impurities bind the carrier, we can think of the impurities as simply adding carriers to the band. So the donor impurities donate free electrons to the conduction band, whereas the acceptor impurities give free holes to the valence band. However, at very low temperature, these carriers get bound back to their respective nuclei so that they can no longer carry electricity, a phenomenon known as *carrier freeze out*.

Note that in the absence of impurities, the Fermi energy (the chemical potential at zero temperature) is in the middle of the band gap. When donor impurities are added, at zero temperature, these states are near the top of the band gap, and are filled. Thus the Fermi energy is moved up to the top of the band gap. On the other hand, when acceptors are added, the acceptor states near the bottom of the band gap are empty. (Remember it is a bound state of a hole to a nucleus!). Thus, the Fermi energy is moved down to the bottom of the band gap.

Optical Effects of Impurities (Redux)

As mentioned previously in section 15.5.4, the presence of impurities in a material can have dramatic effects on its optical properties. There are two main optical effects of impurities. The first effect is that the impurities add charge carriers to an otherwise insulating material – turning an insulator into something that conducts at least somewhat. This obviously can have some important effects on the interaction with light. The second important effect is the introduction of new energy levels within the gap. Whereas before the introduction of impurities, the lowest energy transition that can be made is the full energy of the gap, now one can have optical transitions between impurity states, or from the bands to the impurity states.

¹⁴Note that the large Bohr Radius justifies *post-facto* our use of a continuum approximation for the dielectric constant ϵ_r . On small length scales, the electric field is extremely inhomogeneous due to the microscopic structure of the atoms, but on large enough length scales we can use classical electromagnetism and simply model the material as a medium with a dielectric constant.

¹⁵Because Silicon has an anisotropic band, and therefore an anisotropic mass, the actual formula is more complicated.

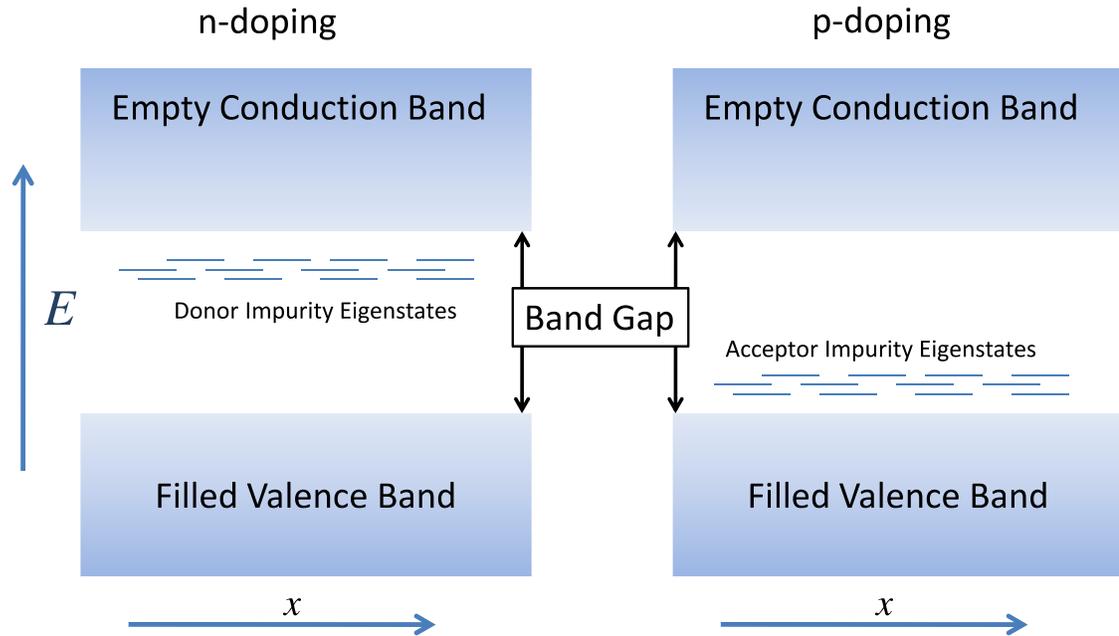


Figure 16.4: Energy Diagram of a Doped Semiconductor (left) with donor impurities (right) with acceptor impurities. The energy eigenstates of the hydrogenic orbitals tied to the impurities are not all the same because each impurity is perturbed by neighbor impurities. At low temperature, the donor impurity eigenstates are filled and the acceptor eigenstates are empty. But with increasing temperature, the electrons in the donor eigenstates are excited into the conduction band and similarly the holes in the acceptor eigenstates are excited into the valence band.

16.3 Statistical Mechanics of Semiconductors

We now use our knowledge of statistical physics to analyze the occupation of the bands at finite temperature.

Imagine a band structure as shown in Fig. 16.5. The minimum energy of the conduction band is defined to be ϵ_c and the maximum energy of the valence band is defined to be ϵ_v . The band gap is correspondingly $E_{gap} = \epsilon_c - \epsilon_v$.

Recall from way back in Eq. 4.10 that the density of states per unit volume for free electrons (in three dimensions with two spin states) is given by

$$g(\epsilon \geq 0) = \frac{(2m)^{3/2}}{2\pi^2\hbar^3} \sqrt{\epsilon}$$

The electrons in our conduction band are exactly like these free electrons, except that (a) the bottom of the band is at energy ϵ_c and (b) they have an effective mass m_e^* . Thus the density

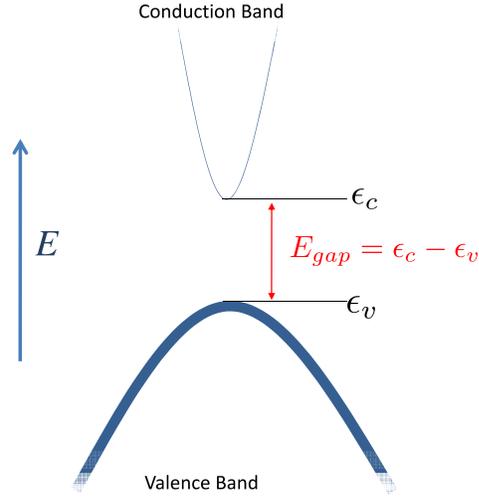


Figure 16.5: A Band Diagram of a Semiconductor.

of states for these electrons near the bottom of the conduction band is given by

$$g_c(\epsilon \geq \epsilon_c) = \frac{(2m_e^*)^{3/2}}{2\pi^2\hbar^3} \sqrt{\epsilon - \epsilon_c}$$

Similarly the density of states for holes near the top of the valence band is given by

$$g_v(\epsilon \leq \epsilon_v) = \frac{(2m_h^*)^{3/2}}{2\pi^2\hbar^3} \sqrt{\epsilon_v - \epsilon}$$

At fixed chemical potential μ the total number of electrons n in the conduction band, as a function of temperature T is thus given by

$$n(T) = \int_{\epsilon_c}^{\infty} d\epsilon g_c(\epsilon) n_F(\beta(\epsilon - \mu)) = \int_{\epsilon_c}^{\infty} d\epsilon \frac{g_c(\epsilon)}{e^{\beta(\epsilon - \mu)} + 1}$$

where n_F is the Fermi occupation factor, and $\beta^{-1} = k_B T$ as usual. If the chemical potential is “well below” the conduction band (i.e., if $\beta(\epsilon - \mu) \gg 1$), then we can approximate

$$\frac{1}{e^{\beta(\epsilon - \mu)} + 1} \approx e^{-\beta(\epsilon - \mu)}$$

In other words, Fermi statistics can be replaced by Boltzmann statistics when the temperature is low enough that the density of electrons in the band is very low. (We have already run into this

principle in section 16.1.1 when we discussed that Drude theory, a classical approach that neglects Fermi statistics, actually works very well for electrons above the band gap in semiconductors!). We thus obtain

$$\begin{aligned} n(T) &\approx \int_{\epsilon_c}^{\infty} d\epsilon g_v(\epsilon) e^{-\beta(\epsilon-\mu)} = \frac{(2m_e^*)^{3/2}}{2\pi^2\hbar^3} \int_{\epsilon_c}^{\infty} d\epsilon (\epsilon - \epsilon_c)^{1/2} e^{-\beta(\epsilon-\mu)} \\ &= \frac{(2m_e^*)^{3/2}}{2\pi^2\hbar^3} e^{\beta(\mu-\epsilon_c)} \int_{\epsilon_c}^{\infty} d\epsilon (\epsilon - \epsilon_c)^{1/2} e^{-\beta(\epsilon-\epsilon_c)} \end{aligned}$$

The last integral is (using $y^2 = x = \epsilon - \epsilon_c$).

$$\int_0^{\infty} dx x^{1/2} e^{-\beta x} = 2 \int_0^{\infty} dy y^2 e^{-\beta y^2} = 2 \frac{d}{d\beta} \int_0^{\infty} e^{-\beta y^2} = \frac{d}{d\beta} \sqrt{\frac{\pi}{\beta}} = \frac{1}{2} \beta^{-3/2} \sqrt{\pi}$$

Thus we obtain the standard expression for the density of electrons in the conduction band

$$n(T) = \frac{1}{4} \left(\frac{2m_e^* k_B T}{\pi \hbar^2} \right)^{3/2} e^{-\beta(\epsilon_c - \mu)} \quad (16.6)$$

Note that this is mainly just exponential activation from the chemical potential to the bottom of the conduction band, with a prefactor which doesn't change too quickly as a function of temperature (obviously the exponential changes very quickly with temperature!).

Quite similarly, we can write the number of holes in the valence band p as a function of temperature¹⁶

$$p(T) = \int_{-\infty}^{\epsilon_v} d\epsilon g_v(\epsilon) \left[1 - \frac{1}{e^{\beta(\epsilon-\mu)} + 1} \right] = \int_{-\infty}^{\epsilon_v} d\epsilon \frac{g_v(\epsilon) e^{\beta(\epsilon-\mu)}}{e^{-\beta(\epsilon-\mu)} + 1}$$

Again, if μ is substantially above the top of the valence band, we have $e^{\beta(\epsilon-\mu)} \ll 1$ so we can replace this by

$$p(T) = \int_{-\infty}^{\epsilon_v} d\epsilon g_v(\epsilon) e^{\beta(\epsilon-\mu)}$$

and the same type of calculation then gives

$$p(T) = \frac{1}{4} \left(\frac{2m_h^* k_B T}{\pi \hbar^2} \right)^{3/2} e^{-\beta(\mu - \epsilon_v)} \quad (16.7)$$

again showing that the holes are activated from the chemical potential *down* into the valence band. (Recall that pushing a hole down into the valence band costs energy!).

¹⁶If the Fermi factor n_F gives the probability that a state is occupied by an electron, then $1 - n_F$ gives the probability that the state is occupied by a hole.

Law of Mass Action

A rather crucial relation is formed by combining Eq. 16.6 with 16.7.

$$\begin{aligned} n(T)p(T) &= \frac{1}{2} \left(\frac{k_B T}{\pi \hbar^2} \right)^3 (m_e^* m_h^*)^{3/2} e^{-\beta(\epsilon_c - \epsilon_v)} \\ &= \frac{1}{2} \left(\frac{k_B T}{\pi \hbar^2} \right)^3 (m_e^* m_h^*)^{3/2} e^{-\beta E_{gap}} \end{aligned} \quad (16.8)$$

where we have used the fact that the gap energy $E_{gap} = \epsilon_c - \epsilon_v$. Eq. 16.8 is sometimes known as the *law of mass action*¹⁷, and it is true independent of doping of the material.

Intrinsic Semiconductors

For an intrinsic (i.e., undoped) semiconductor the number of electrons excited into the conduction band must be equal to the number of holes left behind in the valence band so $p = n$. We can then divide Eq. 16.6 by 16.7 to get

$$1 = \left(\frac{m_e^*}{m_h^*} \right)^{3/2} e^{-\beta(\epsilon_v + \epsilon_c - 2\mu)}$$

Taking log of both sides gives the useful relation

$$\mu = \frac{1}{2}(\epsilon_c + \epsilon_v) + \frac{3}{4}(k_B T) \log(m_h^*/m_e^*) \quad (16.9)$$

Note that at zero temperature, the chemical potential is precisely mid-gap.

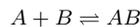
Using either this expression, or by using the law of mass action along with the constraint $n = p$, we can obtain an expression for the *intrinsic* density of carriers in the semiconductor

$$n_{intrinsic} = p_{intrinsic} = \sqrt{np} = \frac{1}{\sqrt{2}} \left(\frac{k_B T}{\pi \hbar^2} \right)^{3/2} (m_e^* m_h^*)^{3/4} e^{-\beta E_{gap}/2}$$

Doped Semiconductors

For doped semiconductors, the law of mass action still holds. If we further assume that the temperature is high enough so that there is no carrier freeze out (i.e, carriers are not bound to

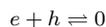
¹⁷The nomenclature here “law of mass action” is a reference to an analogue in chemistry. In chemical reactions we may have an equilibrium between two objects A and B and their compound AB . This is frequently expressed as



There is some chemical equilibrium constant K which gives the ratio of concentrations

$$K = \frac{[A][B]}{[AB]}$$

where $[X]$ is the concentration of species X . The law of mass action states that this constant K remains fixed independent of the individual concentrations. In semiconductor physics it is quite similar, only the “reaction” is



the annihilation of an electron and a hole. So that the product of $[e] = n$ and $[h] = p$ is fixed.

impurities) then we have

$$n - p = (\text{density of donors}) - (\text{density of acceptors})$$

This, along with the law of mass action gives us two equations in two unknowns which can be solved¹⁸. In short, the result is that if we are at a temperature where the undoped intrinsic carrier density is much greater than the dopant density, then the dopants do not matter much, and the chemical potential is roughly midgap as in Eq. 16.9 (This is the *intrinsic* regime). On the other hand, if we are at a temperature where the intrinsic undoped density is much smaller than the dopant density, then the temperature does not matter much and we can think of this as a low temperature situation where the carrier concentration is mainly set by the dopant density (This is the *extrinsic* regime). In the n-doped case, the bottom of the conduction band gets filled with the density of electrons from the donors, and the chemical potential gets shifted up towards the conduction band. Correspondingly, in the p-doped case, holes fill the top of the valence band, and the chemical potential gets shifted down towards the valence band. Note that in this case of strong doping, the majority carrier concentration is obtained just from the doping, whereas the minority carrier concentration — which might be very small — is obtained via law of mass action).

16.4 Summary of Statistical Mechanics of Semiconductors

- Holes are the absence of an electron in the valence band. These have positive charge (electrons have negative charge), and positive effective mass. Energy of a hole gets larger at larger momentum (away from the maximum of the band) as they get pushed down into the valence band. The positive charge of the hole as a charge carrier explains the puzzle of the sign of the Hall coefficient.
- Effective mass of electrons determined by the curvature at the bottom of the conduction band. Effective mass of holes determined by curvature at top of conduction band.
- Mobility of a carrier is $\mu = |e\tau/m^*|$

¹⁸Here is how to solve these two equations. Let

$$D = \text{doping} = n - p = (\text{density of donors}) - (\text{density of acceptors})$$

Let us further assume that $n > p$ so $D > 0$ (we can do the calculation again making the opposite assumption, at the end). Also let

$$I = n_{\text{intrinsic}} = p_{\text{intrinsic}}$$

so that

$$I^2 = \frac{1}{2} \left(\frac{k_B T}{\pi \hbar^2} \right)^3 (m_e^* m_h^*)^{3/2} e^{-\beta E_{\text{gap}}}$$

from the law of mass action. Using $np = I^2$, we can then construct

$$D^2 + 4I^2 = (n - p)^2 + 4np = (n + p)^2$$

So we obtain

$$\begin{aligned} n &= \frac{1}{2} \left(\sqrt{D^2 + 4I^2} + D \right) \\ p &= \frac{1}{2} \left(\sqrt{D^2 + 4I^2} - D \right) \end{aligned}$$

As stated in the main text, if $I \gg D$ then the doping D is not important. On the other hand, if $I \ll D$ then the majority carrier density is determined by the doping only and the thermal factor I is unimportant.

- When very few electrons are excited into the conduction band, or very few holes into the valence band, Boltzmann statistics is a good approximation for Fermi statistics and Drude theory is accurate.
- Electrons or Holes can be excited thermally, or can be added to a system by *doping*. The law of mass action assures that the product np is fixed independent of the amount of doping (only depends on the temperature, the effective masses, and the band gap).
- At very low temperature carriers may freeze-out, binding to the impurity atoms that they came from. However, because the effective Rydberg is very small, carriers are easily ionized into the bands.
- Know how to derive the law of mass action!

References

- Ashcroft and Mermin, chapter 28. A very good discussion of holes and their effective mass is given in chapter 12.
- Rosenberg, chapter 9
- Hook and Hall, 5.1–5.5
- Kittel, chapter 8
- Burns, chapter 10 not including 10.17 and after
- Singleton chapter 5–6

Chapter 17

Semiconductor Devices

The development of semiconductor devices, such as the transistor no doubt changed the world. Every iPad, iPod, iPhone, and iBook literally contains billions of semiconductor transistors. Simple devices, like alarm clocks, TVs, radios, or cars, contain many thousands or even millions of them. It is hard to overstate how much we take these things for granted these days.

This chapter discusses the physics behind some of devices you can make with semiconductors.

17.1 Band Structure Engineering

To make a semiconductor device one must have control over the detailed properties of materials (band gap, doping, etc) and one must be able to assemble together semiconductors with differing such properties.

17.1.1 Designing Band Gaps

A simple example of engineering a device is given by aluminum-galium-arsenide. GaAs is a semiconductor (zincblende structure as in Fig. 13.4) with a direct band gap about $E_{gap,k=0}(\text{GaAs}) = 1.4 \text{ eV}$. AlAs is the same structure except that the Ga has been replaced by Al and the gap¹ at $k = 0$ is about 2.7 eV. One can also produce alloys (mixtures) where some fraction (x) of the Ga has been replaced by Al which we notate as $\text{Al}_x\text{Ga}_{1-x}\text{As}$. To a fairly good approximation the direct band gap just interpolates between the direct band gaps of the pure GaAs and the pure AlAs. Thus we get roughly (for $x < .4$)

$$E_{gap}(x) = (1 - x) 1.4 \text{ eV} + x 2.7 \text{ eV}$$

By producing this type of alloyed structure allows one to obtain any desired band gap in this type of material².

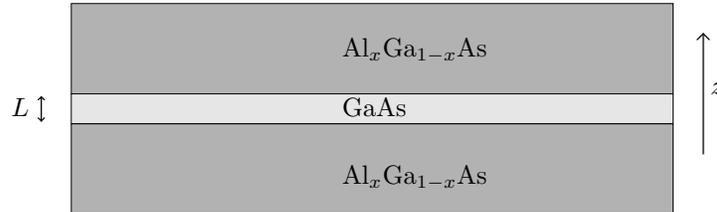
¹AlAs is actually an indirect band gap semiconductor, but for $x < .4$ or so $\text{Al}_x\text{Ga}_{1-x}\text{As}$ is direct band gap as well.

²By alloying the material with arbitrary x , one must accept that the system can no longer be precisely periodic but instead will be some random mixture. It turns out that as long as we are concerned with long wavelength electron waves (i.e, states near the bottom of the conduction band or the top of the valence band) this randomness is very effectively averaged out and we can roughly view the system as being a periodic crystal of As and some average of a $\text{Al}_x\text{Ga}_{1-x}$ atom.. This is known as a “virtual crystal” approximation.

In the context of device physics one might want to build, for example, a laser out of a semiconductor. The lowest energy transition which recombines a hole with an electron is the gap energy (this is the “lasing” energy typically). By tuning the composition of the semiconductor, one can tune the energy of the gap and therefore the optical frequency of the laser.

17.1.2 Non-Homogeneous Band Gaps

By constructing structures where the materials (or the alloying of a material) is a function of position, one can design more complex environments for electrons or holes in a system. Consider for example, the structure shown in the following figure:



Here a layer of GaAs with smaller band gap is inserted between two layers of AlGaAs which has a larger band gap. This structure is known as a “quantum well”. In general a semiconductor made of several varieties of semiconductors is known as a *semiconductor heterostructure*. A band diagram of the quantum well structure as a function of the vertical position z is given in Fig. 17.1. The band gap is lower in the GaAs region than in the AlGaAs region. The changes in band energy can be thought of as a potential that an electron (or hole) would feel. For example, an electron in the valence band can have a lower energy if it is in the quantum well region (the GaAs region) than it can have in the AlGaAs region. An electron in the valence band with low energy will be trapped in this region. Just like a particle in a box, there will be discrete eigenstates of the electron’s motion in the z direction, as shown in the Figure. The situation is similar for holes in the valence band (recall that it requires energy to push a hole down into the valence band), so there will similarly be confined particle in a box states for holes in the quantum well.

17.1.3 Summary of the Examinable Material

- One can tune band gaps by forming an alloy
- Band gaps act as a potential for electrons and holes.

References on Inhomogeneous Semiconductors

There are many good references on semiconductors (See also the material listed below). Almost all of them discuss the p-n junction first (which is nonexaminable for us). I recommend Hook and Hall section 6.6 on the quantum well to cover the above material.

The rest of the material in this chapter is **NOT EXAMINABLE**. But since semiconductors really did change the world, you might be interested in learning it anyway!

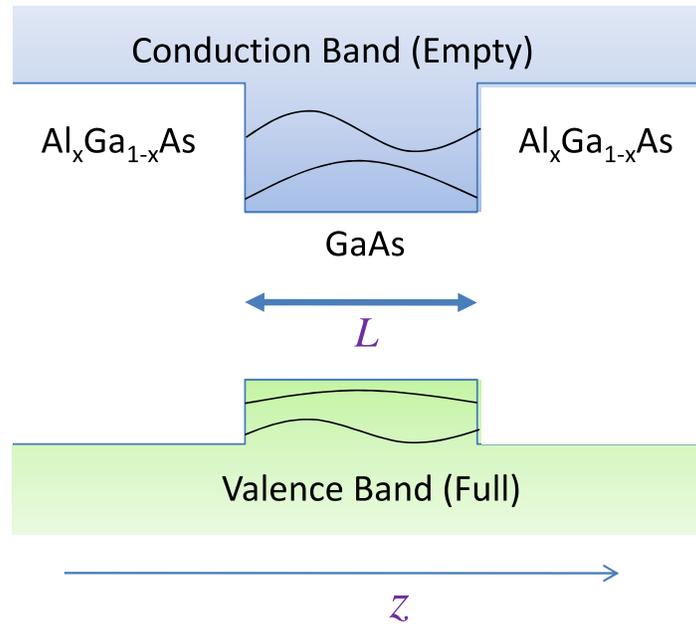


Figure 17.1: Band diagram of a quantum well. A single electron in the conduction band can be trapped in the particle-in-a-box states in the quantum well. Similarly, a hole in the valence band can be trapped in the quantum well.

17.2 p - n Junction

The p - n junction is a junction in a semiconductor between a region of p -doping and a region of n doping. This type of junction has the remarkable property of *rectification*: it will allow current to flow through the junction easily in one direction, but not easily (with very high resistance) in the other direction³.

Consider

OK, I haven't finished this chapter. Cut me some slack, typing these notes was a load of work! Anyway, it is nonexaminable material, so don't worry about it too much.

³The phenomenon of rectification in semiconductors was discovered by Karl Ferdinand Braun way back in 1874, but was not understood in detail until the middle of the next century. This discovery was fundamental to the development of radio technology. Braun was awarded the Nobel Prize in 1909 with Guglielmo Marconi for contributions to wireless telegraphy. Perhaps as important to modern communication, Braun also invented the cathode ray tube (CRT) which formed the display for televisions for many years until the LCD display arrived very recently. (The CRT is known as a "Braun Tube" in many countries).

Part VII

Magnetism and Mean Field Theories

Chapter 18

Magnetic Properties of Atoms: Para- and Dia-Magnetism

The first question one might ask is why we are interested in magnets. While the phenomenon of magnetism was known to the ancients¹, it has only been since the discovery of quantum mechanics that we have come to any understanding of what causes this effect². It may seem like this is a relatively small corner of physics for us to focus so much attention (indeed, several chapters), but we will see that magnetism is a particularly good place to observe the effects of both statistical physics and quantum physics³. As we mentioned in section 15.4, one place where the band theory of electrons fails is in trying to describe magnets. Indeed, this is precisely what makes magnets interesting! In fact, magnetism remains an extremely active area of research in physics (with many many hard and unanswered questions remaining). Much of condensed matter physics continues to use magnetism as a testing ground for understanding complex quantum and statistical physics both theoretically and in the laboratory.

We should emphasize that most magnetic phenomena are caused by the quantum mechanical behavior of *electrons*. While nuclei do have magnetic moments, and therefore can contribute to magnetism, the magnitude of the nuclear moments is (typically) much less than that of electrons.⁴

¹Both the Chinese and the Greeks probably knew about magnetic properties of Fe_3O_4 , or magnetite (also known as loadstone when magnetized), possibly as far back as several thousands of years BCE (with written records existing as far back as 600 years BCE). One legend has it that a shepherd named Magnes, in the province of Magnesia, had the nails in his shoes stuck to a large metallic rock and the scientific phenomenon became named after him.

²Animal magnetism notwithstanding... (that was a joke).

³In fact there is a theorem by Niels Bohr and Hendrika van Leeuwen which shows that any treatment of statistics mechanics without quantum mechanics (i.e., classical statistical mechanics) can never produce a nonzero magnetization.

⁴To understand this, recall that the Bohr magneton, which gives the size of magnetic moments of an electrons is given by $\mu_B = \frac{e\hbar}{2m}$ with m the electron mass. If one were to consider magnetism caused by nuclear moments, the typical moments would be smaller by a ratio of the mass of the electron to the mass of a nucleus (a factor of over 1000). Nonetheless, the magnetism of the nuclei, although small, does exist.

18.1 Basic Definitions of types of Magnetism

Let us first make some definitions. Recall that for a small magnetic field, the magnetization of a system \mathbf{M} (moment per unit volume) is typically related linearly to the applied⁵ magnetic field \mathbf{H} by a (magnetic) susceptibility χ . We write for small fields \mathbf{H} ,

$$\mathbf{M} = \chi\mathbf{H} \quad (18.1)$$

Note that χ is dimensionless. For small susceptibilities (and susceptibilities are almost always small, except in ferromagnets) there is little difference between $\mu_0\mathbf{H}$ and \mathbf{B} (with μ_0 the permeability of free space), so we can also write

$$\mathbf{M} = \chi\mathbf{B}/\mu_0 \quad (18.2)$$

Definition 18.1.1. A *paramagnet* is a material where $\chi > 0$ (i.e., the resulting magnetization is in the same direction as the applied field).

We have run into (Pauli) paramagnetism previously in section 4.3 above. You may also be familiar with the paramagnetism of a free spin (which we will cover again in section 18.4 below). Qualitatively paramagnetism occurs whenever there are moments that can be re-oriented by an applied magnetic field — thus developing magnetization in the direction of the applied field.

Definition 18.1.2. A *diamagnet* is a material where $\chi < 0$ (i.e., the resulting magnetization is in the opposite direction from the applied field).

We will discuss diamagnetism more in section 18.5 below. As we will see, diamagnetism is quite ubiquitous and occurs generically unless it is overwhelmed by other magnetic effects. For example, water, and pretty almost all other biological materials are diamagnetic⁶. Qualitatively we can think of diamagnetism as being similar in spirit to Lenz’s law (part of Faraday’s law) that an induced current always opposes the change causing it. However, the analogy is not precise. If a magnetic field is applied to a loop of wire, current will flow to create a magnetization in the opposite direction. However, in any (nonsuperconducting) loop of wire, the current will eventually decay back to zero and there will be no magnetization remaining. In a diamagnet, in contrast, the magnetization remains so long as the applied magnetic field remains.

For completeness we should also define a ferromagnet — this is what we usually think of as a “magnet” (the thing that holds notes to the fridge).

Definition 18.1.3. A *ferromagnet* is a material where \mathbf{M} can be nonzero, even in the absence of any applied magnetic field⁷

⁵The susceptibility is defined in terms of \mathbf{H} . With a long rod-shaped sample oriented parallel to the applied field, \mathbf{H} is the same outside and inside the sample, and is thus directly controlled by the experimentalist. The susceptibility is defined in terms of this standard configuration. However, more generally, one needs to take care that the internal field \mathbf{B} that any electrons in the sample respond to is related to \mathbf{H} via $\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$.

⁶It is interesting to note that a diamagnet repels the field that creates it, so it is attracted to a magnetic field minimum. Earnshaw’s theorem forbids a local maximum of the \mathbf{B} field in free space, but local minima can exist — and this then allows diamagnets to levitate in free space. In 1997 Andre Geim used this effect to levitate a rather confused frog. This feat earned him a so-called Ig-Nobel prize in 2000 (Ig-Nobel prizes are awarded for research that “cannot or should not be reproduced”). Ten years later he was awarded a real Nobel prize for the discovery of graphene — single layer carbon sheets. This makes him the only person so far to receive both the Ig-Nobel and the real Nobel.

⁷The definition of ferromagnetism given here is a broad definition which would also include ferrimagnets. We will discuss ferrimagnets in section 19.1.3 below and we mention that occasionally people use a more restrictive definition (also commonly used) of ferromagnetism that excludes ferrimagnets. At any rate, the broad definition given here is common.

It is worth already drawing the distinction between *spontaneous* and *non-spontaneous* magnetism. Magnetism is said to be *spontaneous* if it occurs even in the absence of externally applied magnetic field, as is the case for a ferromagnet. The remainder of this chapter will mainly be concerned with non-spontaneous magnetism, and we will return to spontaneous magnetism in chapter 19 below.

It turns out that a lot of the physics of magnetism can be understood by just considering a single atom at a time. This will be the strategy of the current chapter — we will discuss the magnetic behavior of a single atom and only in section 18.6 will consider how the physics changes when we put many atoms together to form a solid. We thus start this discussion by reviewing some atomic physics that you might have learned in prior courses⁸.

18.2 Atomic Physics: Hund's Rules

We start with some of the fundamentals of electrons in an isolated atom. (I.e., we ignore the fact that in materials atoms are not isolated, but are bound to other atoms). For isolated atoms there are a set of rules, known as “Hund's Rules”⁹ which determine how the electrons fill orbitals. Recall from basic quantum mechanics that an electron in an atomic orbital can be labeled by four quantum numbers, $|n, l, l_z, \sigma_z\rangle$, where

$$\begin{aligned} n &= 1, 2, \dots \\ l &= 0, 1, \dots, n - 1 \\ l_z &= -l, \dots, l \\ \sigma_z &= -1/2 \text{ or } +1/2 \end{aligned}$$

Here n is the principle quantum number, l is the angular momentum, l_z is its z -component and σ_z is the z -component of spin¹⁰. Recall that the angular momentum shells with $l = 0, 1, 2, 3$ are sometimes known as s,p,d,f, ... respectively in atomic language. These shells can accommodate 2, 6, 10, 14, ... electrons respectively including both spin states.

When we consider multiple electrons in one atom, we need to decide which orbitals are filled and which ones are empty. The first rule is known as the Aufbau principle¹¹, which many people

⁸You *should* have learned this in prior courses. But if you not, it is probably not your fault! This material is rarely taught in physics courses these days, even though it really should be. Much of this material is actually taught in chemistry courses instead!

⁹Friedrich Hermann Hund was an important physicist and chemist whose work on atomic structure began in the very early days of quantum mechanics — he wrote down Hund's rules in 1925. He is also credited with being one of the inventors of molecular orbital theory which we met in chapter 5.3.2 above. In fact, molecular orbital theory is sometimes known as Hund-Mulliken Molecular Orbital theory. Mulliken thanked Hund heavily in his Nobel Prize acceptance speech (but Hund did not share the prize). Hund died in 1997 at the age of 101. The word “Hund” means “Dog” in German.

¹⁰You probably discussed these quantum numbers in reference to the eigenstates of a Hydrogen atom. The orbitals of any atom can be labeled similarly.

¹¹Aufbau means “construction” or “building up” in German.

think of as Hund's 0th rule

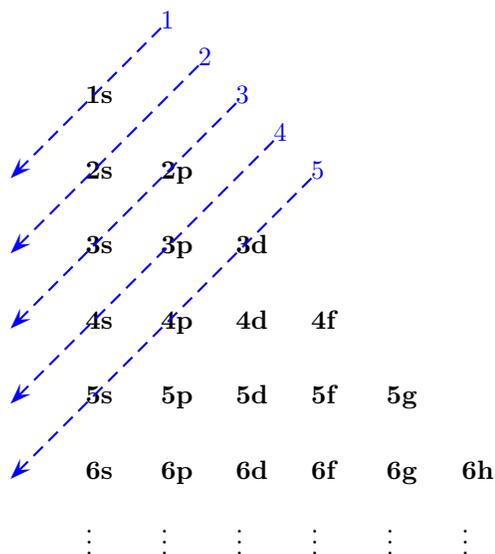
Aufbau Principle (paraphrased): Shells should be filled starting with the lowest available energy state. An entire shell is filled before another shell is started¹².

(Madelung Rule): The energy ordering is from lowest value of $n+l$ to the largest; and when two shells have the same value of $n+l$, fill the one with the smaller n first.¹³

This ordering rule means that shells should be filled in the order¹⁴

$$1s, 2s, 2p, 3s, 3p, 4s, 3d, 4p, 5s, 4d, 5p, 6s, 4f, \dots$$

A simple mnemonic for this order can be constructed by drawing the following simple diagram:



So for example, let us consider an isolated nitrogen atom which has atomic number 7 (i.e., 7 electrons). Nitrogen has a filled $1s$ shell (containing 2 electrons, one spin up, one spin down), has a filled $2s$ shell (containing 2 electrons, one spin up, one spin down), and has three remaining electrons in the $2p$ shell. In atomic notation we would write this as $1s^2 2s^2 2p^3$.

To take a more complicated example, consider the atom praseodymium (Pr) which is a

¹²It is important to realize that a given orbital is different in different atoms. For example, the $2s$ orbital in a nitrogen atom is different from the $2s$ orbital in an iron atom. The reason for this is that the charge of the nucleus is different *and also* that one must account for the interaction of an electron in an orbital with all of the other electrons in that atom.

¹³Depending on your country of origin, the Madelung rule might instead be known as Klechkovsky's rule.

¹⁴You may find it surprising that shells are filled in this order, being that for a simple hydrogen atom orbital energies increase with n and are independent of l . However, in any atom other than hydrogen, we must also consider interaction of each electron with all of the other electrons. Treating this effect in detail is quite complex, so it is probably best to consider the this ordering (Madelung) rule to be simply empirical. Nonetheless, various approximation methods have been able to give some insight. Typical approximation schemes replace the Coulomb potential of the nucleus with some screened potential which represents the charge both of the nucleus and of all the other electrons (essentially the effective charge of the nucleus is reduced if the electron is at a radius where other electrons can get between it and the nucleus). Note in particular that once we changes the potential from the Coulomb $1/r$ form, we immediately break the energy degeneracy between different l states.

rare earth element with atomic number 59. Following the Madelung rule, we obtain an atomic¹⁵ configuration $1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^{10} 4p^6 5s^2 4d^{10} 5p^6 6s^2 4f^3$. Note that the “exponents” properly add up to 59.

There are a few atoms that violate this ordering (Madelung) rule. One example is copper which typically fills the 3d shell by “borrowing” an electron from the (putatively lower energy) 4s shell. Also, when an atom is part of a molecule or is in a solid, the ordering may change a little as well. However, the general trend given by this rule is rather robust.

This shell filling sequence is, in fact, the rule which defines the overall structure of the periodic table with each “block” of the periodic table representing the filling of some particular shell. For example, the first line of the periodic table has the elements H and He, which have atomic fillings $1s^x$ with $x = 1, 2$ respectively (and the 1s shell holds at most 2 electrons). The left of the second line of the table contains Li and Be which have atomic fillings $1s^2 2s^x$ with $x = 1, 2$ respectively. The right of the 2nd line of the table shows B, N, C, O, F, Ne which have atomic fillings $1s^2 2s^2 2p^x$ with $x = 1 \dots 6$ and recall that the 2p shell can hold at most 6 electrons. One can continue and reconstruct the entire periodic table this way!

In cases when shells are partially filled (which in fact includes most elements of the periodic table) we next want to describe which of the available orbitals are filled in these shells and which spin states are filled. In particular we want to know what whether these electrons will have a net magnetic moment. Hund's rules are constructed precisely to answer this questions.

Perhaps the simplest way to illustrate these rules is to consider an explicit example. Here we will again consider the atom praseodymium. As mentioned above, this element in atomic form has three electrons in its outer-most shell, which is an f -shell, meaning it has angular momentum $l = 3$, and therefore 7 possible values of l_z , and of course 2 possible values of the spin for each electron. So where in these possible orbital/spin states do we put the three electrons?

Hund's First Rule (paraphrased): Electrons try to align their spins.

Given this rule, we know that the three valence electrons in Pr will have their spins point in the same direction, thus giving us a total spin-angular momentum $S = 3/2$ from the three $S = 1/2$ spins. So locally (meaning on the same atom), the three electron spins behave ferromagnetically — they all align¹⁶. The reason for this alignment will be discussed below in section 18.2.1, but in short, it is a result of the Coulomb interaction between electrons (and between the electrons and the nucleus) — the Coulomb energy is lower when the electron spins align.

We now have to decide which orbital states to put the electrons in.

Hund's Second Rule(paraphrased): Electrons try to maximize their total orbital angular momentum, consistent with Hund's first rule.

For the case of Pr, we fill the $l_z = 3$ and $l_z = 2$ and $l_z = 1$ states to make the maximum possible total $L_z = 6$ (this gives $L = 6$, and by rotational invariance we can point \mathbf{L} in any direction

¹⁵This tediously long atomic configuration can be abbreviated as $[\text{Xe}]6s^2 4f^3$ where $[\text{Xe}]$ represents the atomic configuration of Xenon, which, being a noble gas is made of entirely filled shells.

¹⁶We would not call this a true ferromagnet since we are talking about a single atom here, not a macroscopic material!

equally well). Thus, we have a picture as follows

$$\begin{array}{ccccccc}
 & \text{---} & \text{---} & \text{---} & \text{---} & \uparrow & \uparrow & \uparrow \\
 l_z = & -3 & -2 & -1 & 0 & 1 & 2 & 3
 \end{array}$$

We have put the spins as far as possible to the right to maximize L_z (Hund's 2nd rule) and we have aligned all the spins (Hund's 1st rule). Note that we could not have put both electrons in the same orbital, since they have to be spin-aligned and we must obey the Pauli principle. Again the rule of maximizing orbital angular momentum is driven by the physics of Coulomb interaction (as we will discuss briefly below in section 18.2.1).

At this point we have $S = 3/2$ and $L = 6$, but we still need to think about how the spin and orbital angular momenta align with respect to each other.

Hund's Third Rule(paraphrased): Given Hund's first and second rules, the orbital and spin angular momentum either align or antialign, so that the total angular momentum is $J = |L \pm S|$ with the sign being determined by whether the shell of orbitals is more than half filled (+) or less than half filled (-).

The reason for this rule is not interaction physics, but is spin-orbit coupling. The Hamiltonian will typically have a spin-orbit term $\alpha \mathbf{l} \cdot \boldsymbol{\sigma}$, and the sign of α determines how the spin and orbit align to minimize the energy.¹⁷ Thus for the case of Pr, where $L = 6$ and $S = 3/2$ and the shell is less than half filled, we have total angular momentum $J = L - S = 9/2$.

One should be warned that people frequently refer to J as being the "spin" of the atom. This is a colloquial use which is very persistent but imprecise. More correctly, J is the total angular momentum of the electrons in the atom, whereas S is the spin component of J .

18.2.1 Why Moments Align

We now return, as promised above, to discuss roughly why Hund's rules work — in particular we want to know why magnetic moments (real spin moments or orbital moments) like to align with each other. This section will be only qualitative, but should give at least a rough idea of the right physics.

Let us first focus on Hund's first rule and ask why spins like to align. First of all, we emphasize that it has nothing to do with magnetic dipole interactions. While the magnetic dipoles of the spins do interact with each other, when dipole moments are on the order of the Bohr magneton, this energy scale becomes tiny — way too small to matter for anything interesting. Instead, the alignment comes from the Coulomb interaction energy. To see how this works, let us consider a wavefunction for two electrons on an atom.

¹⁷The fact that the sign switches at half filling does not signal a change in the sign of the underlying α (which is always positive) but rather is a feature of careful bookkeeping. So long as the shell remains less than half full, all of the spins are aligned in which case we have $\sum_i \mathbf{l}_i \cdot \boldsymbol{\sigma}_i = \mathbf{S} \cdot \mathbf{L}$ thus always favoring \mathbf{L} counter-aligned with \mathbf{S} . When the shell is half filled $L = 0$. When we add one more spin to a half filled shell, this spin must counter-align with the many spins that comprise the half-filled shell due to the Pauli exclusion principle. The spin orbit coupling $\mathbf{l}_i \cdot \boldsymbol{\sigma}_i$ then makes this additional spin want to counter-align with its own orbital angular momentum \mathbf{l}_i , which is equal to the total orbital angular momentum \mathbf{L} since the half full shell has $L = 0$. This means that the orbital angular momentum is now *aligned* with the net spin, since most of the net spin is made up of the spins comprising the half-filled shell and are counter-aligned with the spin of the electron which has been added.

Naive Argument

The overall wavefunction must be antisymmetric by Pauli's exclusion principle. We can generally write

$$\Psi(\mathbf{r}_1, \sigma_1; \mathbf{r}_2, \sigma_2) = \psi_{orbital}(\mathbf{r}_1, \mathbf{r}_2) \chi_{spin}(\sigma_1, \sigma_2)$$

where \mathbf{r}_i are the particles' positions and σ_i are their spin. Now, if the two spins are aligned, say both are spin-up (i.e., $\chi_{spin}(\uparrow, \uparrow) = 1$ and $\chi_{spin} = 0$ for other spin configurations) then the spin wavefunction is symmetric and the spatial wavefunction $\psi_{orbital}$ must be antisymmetric. As a result we have

$$\lim_{\mathbf{r}_1 \rightarrow \mathbf{r}_2} \psi_{orbital}(\mathbf{r}_1, \mathbf{r}_2) \rightarrow 0$$

So electrons with aligned spins cannot get close to each other, thus reducing the Coulomb energy of the system.

The argument we have just given is frequently stated in textbooks. Unfortunately, it is not the whole story.

More Correct

In fact it turns out that the crucial Coulomb interaction is that between the electron and the nucleus. Consider the case where there are two electrons and a nucleus as shown in Fig. 18.1. What we see from this figure is that the positive charge of the nucleus seen by one electron is screened by the negative charge of the other electron. This screening reduces the binding energy of the electrons to the nucleus. However, when the two spins are aligned, the electrons repel each other and therefore screen the nucleus less effectively. In this case, the electrons see the full charge of the nucleus and bind more strongly, thus lowering their energies.

Another way of understanding this is to realize that when the spins are not aligned, sometimes one electron gets between the other electron and the nucleus — thereby reducing the effective charge seen by the outer electron, reducing the binding energy, and increasing the total energy of the atom. However, when the electrons are spin aligned, the Pauli principle largely prevents this configuration from occurring, thereby lowering the total energy of the system.

Hund's second rule is driven by very similar considerations. When two electrons take states which maximize their total orbital angular momentum, they are more likely to be found on opposite sides of the nucleus. Thus the electrons see the nucleus fully unscreened so that the binding energy is increased and the energy is lowered.

One must be somewhat careful with these types of arguments however — particularly when they are applied to molecules instead of atoms. In the case of a diatomic molecule, say H_2 , we have two electrons and two nuclei. While the screening effect discussed above still occurs, and tries to align the electrons, it is somewhat less effective than for two electrons on a single atom — since most of the time the two electrons are near opposite nuclei anyway. Furthermore, there is a competing effect that tends to make the electrons want to anti-align. As we discussed in section 5.3.1 when we discussed covalent bonding, we can think of the two nuclei as being a square well (See Fig. 5.4), and the bonding is really a particle-in-a-box problem. There is some lowest energy (symmetric) wavefunction in this large two-atom box, and the lowest energy state of two electrons would be to have the two spins anti-aligned so that both electrons can go in the same low energy spatial wavefunction. It can thus be quite difficult to determine whether electrons on neighboring atoms want to be aligned or anti-aligned. Generally either behavior is possible. We will discuss this

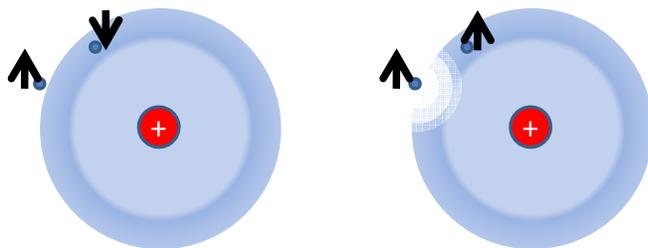


Figure 18.1: Why Aligned Spins Have Lower Energy (Hund’s First Rule). In this figure, the wavefunction is depicted for one of the electrons whereas the other electron (the one further left) is depicted as having fixed position. When the two electrons have opposite spin, the effective charge of the nucleus seen by the fixed electron is reduced by the screening provided by the other electron (left figure) . However, when the spins are aligned, the two electrons cannot come close to each other (right figure) and the fixed electron sees the full charge of the nucleus. As such, the binding of the fixed electron to the nucleus is stronger in the case where the two electrons are spin aligned, therefore it is a lower energy configuration.

much further below in chapter 22. The energy difference between having the spins on two atoms aligned versus anti-aligned is usually known as the *exchange interaction* or *exchange energy*.¹⁸

18.3 Coupling of Electrons in Atoms to an External Field

Having discussed how electron moments (orbital or spin) can align with each other, we now turn to discuss how the electrons in atoms couple to an external magnetic field.

In the absence of a magnetic field, the Hamiltonian for an electron in an atom is of the usual

¹⁸The astute reader will recall that atomic physicists use the word “exchange” to refer to what we called the hopping matrix element (see footnote 14 in section 5.3.2) which “exchanged” an electron from one orbital to another. In fact the current name is very closely related. Let us attempt a very simple calculation of the difference in energy between two electrons having their spins aligned and two electrons having their spins antialigned. Suppose we have two electrons on two different orbitals which we will call A and B . We write a general wavefunction as $\psi = \psi_{spatial}\chi_{spin}$ and overall the wavefunction must be antisymmetric. If we choose the spins to be aligned (a triplet, therefore antisymmetric), then the spatial wavefunction must be symmetric, which we can write as $|AB\rangle + |BA\rangle$. On the other hand if we choose the spins to be anti-aligned (a singlet, therefore symmetric) then the spatial wavefunction must be antisymmetric $|AB\rangle - |BA\rangle$. When we add Coulomb interaction, the energy difference between the singlet and triplet is proportional to the cross term $\langle AB|V|BA\rangle$. In this matrix element the two electrons have “exchanged” place. Hence the name.

form¹⁹

$$\mathcal{H}_0 = \frac{\mathbf{p}^2}{2m} + V(\mathbf{r})$$

where V is the electrostatic potential from the nucleus (and perhaps from the other electrons as well). Now consider adding an external magnetic field. Recall that the Hamiltonian for a charged particle in a magnetic field \mathbf{B} takes the minimal coupling form²⁰

$$\mathcal{H} = \frac{(\mathbf{p} + e\mathbf{A})^2}{2m} + g\mu_B\mathbf{B} \cdot \boldsymbol{\sigma} + V(\mathbf{r})$$

where $-e$ is the charge of the particle (the electron), $\boldsymbol{\sigma}$ is the electron spin, g is the electron g -factor (approximately 2), $\mu_B = e\hbar/(2m)$ is the Bohr magneton, and \mathbf{A} is the vector potential. For a uniform magnetic field, we may take $\mathbf{A} = \frac{1}{2}\mathbf{B} \times \mathbf{r}$ such that $\nabla \times \mathbf{A} = \mathbf{B}$. We then have²¹

$$\mathcal{H} = \frac{\mathbf{p}^2}{2m} + V(\mathbf{r}) + \frac{e}{2m}\mathbf{p} \cdot (\mathbf{B} \times \mathbf{r}) + \frac{e^2}{2m} \frac{1}{4} |\mathbf{B} \times \mathbf{r}|^2 + g\mu_B\mathbf{B} \cdot \boldsymbol{\sigma} \quad (18.3)$$

The first two terms in this equation comprise the Hamiltonian \mathcal{H}_0 in the absence of the applied magnetic field. The next term can be rewritten as

$$\frac{e}{2m}\mathbf{p} \cdot (\mathbf{B} \times \mathbf{r}) = \frac{e}{2m}\mathbf{B} \cdot (\mathbf{r} \times \mathbf{p}) = \mu_B\mathbf{B} \cdot \mathbf{l} \quad (18.4)$$

where $\hbar\mathbf{l} = \mathbf{r} \times \mathbf{p}$ is the orbital angular momentum of the electron. This can then be combined with the so-called Zeeman term $g\mu_B\mathbf{B} \cdot \boldsymbol{\sigma}$ to give

$$\mathcal{H} = \mathcal{H}_0 + \mu_B\mathbf{B} \cdot (\mathbf{l} + g\boldsymbol{\sigma}) + \frac{e^2}{2m} \frac{1}{4} |\mathbf{B} \times \mathbf{r}|^2 \quad (18.5)$$

The second term on the right of this equation, known sometimes as the *paramagnetic* term, is clearly just the coupling of the external field to the total magnetic moment of the electron (both orbital and spin). Note that when a \mathbf{B} -field is applied, these moments aligns with the \mathbf{B} -field (meaning that \mathbf{l} and $\boldsymbol{\sigma}$ anti-align with \mathbf{B}) such that the energy is lowered by the application of the field²². As a result a moment is created in the same direction as the applied field and this term results in paramagnetism.

The final term of Eq. 18.5 is known as the *diamagnetic term* of the Hamiltonian, and will be responsible for the effect of diamagnetism. Since this term is quadratic in \mathbf{B} it will always cause an *increase* in the total energy of the atom when the magnetic field is applied, and hence has the opposite effect from that of the above considered paramagnetic term.

These two terms of the Hamiltonian are the ones responsible for both the paramagnetic and diamagnetic response of atoms to external magnetic fields. We will treat them each in turn in the next two sections. Keep in mind that at this point we are still considering the magnetic response of a single atom!

¹⁹Again, whenever we discuss magnetism it is typical to use \mathcal{H} for the Hamiltonian so as not to confuse it with the magnetic field strength $H = B/\mu_0$.

²⁰Recall that minimal coupling requires $\mathbf{p} \rightarrow \mathbf{p} - q\mathbf{A}$ where q is the charge of the particle. Here our particle has charge $q = -e$. The negative charge also is responsible for the fact that the electron spin magnetic moment is anti-aligned with its spin. Hence it is lower energy to have the spin point opposite the applied magnetic field (hence the positive sign of the so-called Zeeman term $g\mu_B\mathbf{B} \cdot \boldsymbol{\sigma}$). Blame Ben Franklin. (See footnote 13 of section 4.3).

²¹Note that while p_i does not commute with r_i , it does commute with r_j for $j \neq i$, so there is no ordering problem between \mathbf{p} and $\mathbf{B} \times \mathbf{r}$

²²If the sign of the magnetic moment confuses you, it is good to remember that moment is always $-\partial F/\partial B$, and at zero temperature the free energy is just the energy.

18.4 Free Spin (Curie or Langevin) Paramagnetism

We will start by considering the effect of the paramagnetic term of Eq. 18.5. We assume that the unperturbed Hamiltonian \mathcal{H}_0 has been solved and we need not pay attention to this part of the Hamiltonian — we are only concerned with the reorientation of a spin $\boldsymbol{\sigma}$ or an orbital angular momentum \mathbf{I} of an electron. At this point we also disregard the diamagnetic term of the Hamiltonian as its effect is generally weaker than that of the paramagnetic term.

Free Spin 1/2

As a review let us consider a simpler case that you are probably familiar with from your statistical physics course: a free spin-1/2. The Hamiltonian, you recall, of a single spin-1/2 is given by

$$\mathcal{H} = g\mu_B \mathbf{B} \cdot \boldsymbol{\sigma} \quad (18.6)$$

with g the g-factor of the spin which we set to be 2, and $\mu_B = e\hbar/(2m)$ is the Bohr magneton. We can think of this as being a simplified version of the above paramagnetic term of Eq. 18.5, for a single free electron where we ignore the orbital moment. The eigenstates of $\mathbf{B} \cdot \boldsymbol{\sigma}$ are $\pm B/2$ so we have a partition function

$$Z = e^{-\beta\mu_B B} + e^{\beta\mu_B B} \quad (18.7)$$

and a corresponding free energy $F = -k_B T \log Z$ giving us a magnetic moment (per spin) of

$$\text{moment} = -\frac{\partial F}{\partial B} = \mu_B \tanh(\beta\mu_B B) \quad (18.8)$$

If we have many such atoms together in a volume, we can define the magnetization M to be the magnetic moment per unit volume. Then, at small field (expanding the tanh for small argument) we obtain a susceptibility of

$$\chi = \lim_{H \rightarrow 0} \frac{\partial M}{\partial H} = \frac{n\mu_0\mu_B^2}{k_b T} \quad (18.9)$$

where n is the number of spins per unit volume (and we have used $B \approx \mu_0 H$ with μ_0 the permeability of free space). Expression 18.9 is known as the “Curie Law”²³ susceptibility (Actually any susceptibility of the form $\chi \sim C/(k_b T)$ for any constant C is known as Curie law), and paramagnetism involving free spins like this is often called Curie paramagnetism or Langevin²⁴ paramagnetism.

Free Spin J

The actual paramagnetic term in the Hamiltonian will typically be more complicated than our simple spin-1/2 model, Eq. 18.6. Instead, examining Eq. 18.5 and generalizing to multiple electrons

²³Named after Pierre Curie. Pierre’s work on magnetism was well before he married his mega-brilliant wife Marie Sklodowska Curie. She won one physics Nobel with Pierre, and then another one in chemistry after he died. Half-way between the two prizes, Pierre was killed when he was run over by a horse-drawn vehicle while crossing the street (Be careful!).

²⁴Paul Langevin was Pierre Curie’s student. He is well known for many important scientific discoveries. He is also well known for creating quite the scandal by having an affair with Marie Curie a few years after her husband’s death (Langevin was married at the time). Although the affair quickly ended, ironically, the grandson of Langevin married the granddaughter of Curie and they had a son — all three of them are physicists.

in an atom, we expect to need to consider a Hamiltonian of the form

$$\mathcal{H} = \mu_B \mathbf{B} \cdot (\mathbf{L} + g\mathbf{S}) \quad (18.10)$$

where \mathbf{L} and \mathbf{S} are the orbital and spin components of all of the electrons in the atom put together. Recall that Hund's rules tell us the value of L , S , and J . The form of Eq. 18.10 looks a bit inconvenient, since Hund's third rule tells us not about $\mathbf{L} + g\mathbf{S}$ but rather tells us about $\mathbf{J} = \mathbf{L} + \mathbf{S}$. Fortunately, for the type of matrix elements we are concerned with (reorientations of \mathbf{J} without changing the value of J , S or L which are dictated by Hund's rules) the above Hamiltonian Eq. 18.10 turns out to be precisely equivalent to

$$\mathcal{H} = \tilde{g}\mu_B \mathbf{B} \cdot \mathbf{J} \quad (18.11)$$

where \tilde{g} is an effective g -factor given by²⁵

$$\tilde{g} = \frac{1}{2}(g+1) + \frac{1}{2}(g-1) \left[\frac{S(S+1) - L(L+1)}{J(J+1)} \right]$$

From our new Hamiltonian, it is easy enough to construct the partition function

$$Z = \sum_{J_z=-J}^J e^{\beta \tilde{g}\mu_B B J_z} \quad (18.12)$$

Analogous to the spin-1/2 case above one can differentiate to obtain the moment as a function of temperature. If one considers a density n of these atoms, one can then determine the magnetization and the susceptibility (this is assigned as an "Additional Problem" for those who are interested). The result, of the Curie form, is that the susceptibility per unit volume is given by

$$\chi = \frac{n\mu_0(\tilde{g}\mu_B)^2 J(J+1)}{3 k_B T}$$

(Compare Eq. 18.9)

Note that Curie law susceptibility always *diverges* at low temperature²⁶. If this term is nonzero (i.e., if J is nonzero) then the Curie paramagnetism is dominant compared to any other type of paramagnetism or diamagnetism²⁷.

²⁵You probably do not need to memorize this formula for this course, although you might have to know it for atomic physics! The derivation of this formula is not difficult though. We are concerned in determining matrix elements of $\mathbf{B} \cdot (\mathbf{L} + g\mathbf{S})$ between different J_z states. To do this we write

$$\mathbf{B} \cdot (\mathbf{L} + g\mathbf{S}) = \mathbf{B} \cdot \mathbf{J} \left[\frac{\mathbf{L} \cdot \mathbf{J}}{|\mathbf{J}|^2} + g \frac{\mathbf{S} \cdot \mathbf{J}}{|\mathbf{J}|^2} \right]$$

The final bracket turns out to be just a number, which we evaluate by rewriting it as

$$\left[\frac{|\mathbf{J}|^2 + |\mathbf{L}|^2 - |\mathbf{J} - \mathbf{L}|^2}{2|\mathbf{J}|^2} \right] + g \left[\frac{|\mathbf{J}|^2 + |\mathbf{S}|^2 - |\mathbf{J} - \mathbf{S}|^2}{2|\mathbf{J}|^2} \right]$$

Finally replacing $\mathbf{J} - \mathbf{L} = \mathbf{S}$ and $\mathbf{J} - \mathbf{S} = \mathbf{L}$ then substituting in $|\mathbf{J}|^2 = J(J+1)$ and $|\mathbf{S}|^2 = S(S+1)$ and $|\mathbf{L}|^2 = L(L+1)$, with a small bit of algebra gives the desired result.

²⁶The current calculation is a finite temperature thermodynamic calculation resulting in divergent susceptibility at zero temperature. In the next few sections we will study Larmor and Landau Diamagnetism as well as Pauli and Van-Vleck Paramagnetism. All of these calculations will be zero temperature quantum calculations and will always give much smaller finite susceptibilities.

²⁷Not including superconductivity.

Aside: From Eqs. 18.7 or 18.12 we notice that the partition function of a free spin is only a function of the dimensionless ratio $\mu_B B / (k_b T)$. From this we can derive that the entropy S is also a function only of the same dimensionless ratio. Let us imagine now we have a system of free spins at magnetic field B and temperature T , and we thermally isolate it from the environment. If we adiabatically reduce B , then since S must stay fixed, the temperature must drop proportionally to the reduction in B . This is the principle of the adiabatic demagnetization refrigerator.²⁸

18.5 Larmor Diamagnetism

Since Curie paramagnetism is dominant whenever $J \neq 0$, the only time we can possibly observe diamagnetism is if an atom has $J = 0$. A classic situation in which this occurs is for atoms with filled shell configurations, like the noble gases where $L = S = J = 0$. Another possibility is that $J = 0$ even though $L = S$ is nonzero (one can use Hund's rules to show that this occurs if a shell has one electron fewer than being half filled). In either case, the paramagnetic term of Eq. 18.5 has zero expectation and the term can be mostly ignored²⁹. We thus need to consider the effect of the final term in Eq. 18.5, the diamagnetic term.

If we imagine that \mathbf{B} is applied in the \hat{z} direction, the expectation of the diamagnetic term of the Hamiltonian (Eq. 18.5) can be written as

$$\delta E = \frac{e^2}{8m} \langle |\mathbf{B} \times \mathbf{r}|^2 \rangle = \frac{e^2 B^2}{8m} \langle x^2 + y^2 \rangle$$

Using the fact that the atom is rotationally symmetric, we can write

$$\langle x^2 + y^2 \rangle = \frac{2}{3} \langle x^2 + y^2 + z^2 \rangle = \frac{2}{3} \langle r^2 \rangle$$

Thus we have

$$\delta E = \frac{e^2 B^2}{12m} \langle r^2 \rangle$$

Thus the magnetic moment per electron is

$$\text{moment} = -\frac{dE}{dB} = -\left[\frac{e^2}{6m} \langle r^2 \rangle \right] B$$

²⁸Very low temperature adiabatic demagnetization refrigerators usually rely on using *nuclear* moments rather than electronic moments. The reason for this is that the (required) approximation of spins being independent holds down to much lower temperature for nuclei, which are typically quite decoupled from their neighbors. Achieving nuclear temperatures below 1 μ K is possible with technique.

²⁹Actually, to be more precise, even though $\langle \mathbf{J} \rangle$ may be zero, the paramagnetic term in Eq. 18.5 may be important in second order perturbation theory. At second order, the energy of the system will be corrected by a term proportional to

$$\delta E_0 \sim + \sum_{p>0} \frac{|\langle p | \mathbf{B} \cdot (\mathbf{L} + g\mathbf{S}) | 0 \rangle|^2}{E_0 - E_p}$$

This contribution need not vanish. It is largest when there is a low energy excitation E_p so the denominator can be small. Since this energy *decreases* with increasing B , this term is paramagnetic. At any rate, this contribution is often important in the cases where $J = 0$ but L and S are individually nonzero — as this usually implies there is a low energy excitation that can occur by misorienting \mathbf{L} and \mathbf{S} with respect to each other thus violating Hund's 3rd rule only. However, for atoms like noble gases, where \mathbf{L} and \mathbf{S} are individually zero, then there are no low energy excitations and this contribution is negligible. This type of paramagnetism is known as *Van Vleck paramagnetism* after the Nobel Laureate J. H. Van Vleck who was a professor at Baliol college Oxford in 1961–1962 but spent most of his later professional life at Harvard.

Assuming that there is a density ρ of such electrons in a system, we can then write the susceptibility as

$$\chi = -\frac{\rho e^2 \mu_0 \langle r^2 \rangle}{6m} \quad (18.13)$$

This result, Eq. 18.13, is known as *Larmor Diamagnetism*.³⁰ For most atoms, $\langle r^2 \rangle$ is on the order of a few Bohr radii squared. In fact, the same expression can sometimes be applied for large conductive molecules if the electrons can freely travel the length of the molecule — by taking $\langle r^2 \rangle$ to be the radius squared of the molecule instead of that of the atom.

18.6 Atoms in Solids

Up to this point, we have always been considering the magnetism (paramagnetism or diamagnetism) of a single isolated atom. Although the atomic picture gives a very good idea of how magnetism occurs, the situation in solids can be somewhat different. As we have discussed in chapters 14 and 15 when atoms are put together the electronic band structure defines the physics of the material — we cannot usually think of atoms as being isolated from each other. We thus must think a bit more carefully about how our above atomic calculations may or may not apply to real materials.

18.6.1 Pauli Paramagnetism in Metals

Recall that in section 4.3 we calculated the susceptibility of the free Fermi gas. We found

$$\chi_{Pauli} = \mu_0 \mu_B^2 g(E_F) \quad (18.14)$$

with $g(E_F)$ the density of states at the Fermi surface. We might expect that such an expression would hold for metals with nontrivial band structure — only the density of states would need to be modified. Indeed, such an expression holds fairly well for simple metals such as Li or Na.

Note that the susceptibility, per spin, of a Fermi gas (Eq. 18.14) is smaller than the susceptibility of a free spin (Eq. 18.9) by roughly a factor of T/E_F (This can be proven using Eq. 4.11 for a free electron gas). We should be familiar this idea, that due to the Pauli exclusion principle, only the small fraction of spins near the Fermi surface can be flipped over, therefore giving a small susceptibility.

18.6.2 Diamagnetism in Solids

Our above calculation of Larmor diamagnetism was applied to isolated atoms each having $J = L = S = 0$, such as noble gas atoms. At low temperature noble gas atoms form very weakly bonded crystals and the same calculation continues to apply (with the exception of the case of helium which does not crystalize but rather forms a superfluid at low temperature). To apply the above result Eq. 18.13 to a noble gas crystal, one simply sets the density of electrons ρ to be equal to the density of atoms n times the number of electrons per atom (the atomic number) Z . Thus for noble gas atoms we obtain

$$\chi_{Larmor} = -\frac{Zne^2 \mu_0 \langle r^2 \rangle}{6m} \quad (18.15)$$

³⁰Joseph Larmor was a rather important physicist in the late 1800s. Among other things, he published the Lorentz transformations for time dilation and length contraction two years before Lorentz, and seven years before Einstein. However, he insisted on the aether, and rejected relativity at least until 1927 (maybe longer).

where $\langle r^2 \rangle$ is set by the atomic radius.

In fact, for any material the diamagnetic term of the Hamiltonian (the coupling of the orbital motion to the magnetic field) will result in some amount of diamagnetism. To account for the diamagnetism of electrons in core orbitals, Eq. 18.15 is usually fairly accurate. For the conduction electrons in a metal, however, a much more complicated calculation gives the so-called Landau-diamagnetism (See footnote 12 of chapter 3)

$$\chi_{Landau} = -\frac{1}{3}\chi_{Pauli}$$

which combines with the Pauli paramagnetism to reduce the total paramagnetism of the conduction electrons by 1/3.

If one considers, for example, a metal like copper, one might be tempted to conclude that it should be a paramagnet, due to the above described Pauli paramagnetism (corrected by the Landau effect). However, copper is actually a diamagnet! The reason for this is that the core electrons in copper have enough Larmor diamagnetism to overwhelm the Pauli paramagnetism of the conduction electrons! In fact, Larmor diamagnetism is often strong enough to overwhelm Pauli paramagnetism in metals (this is particularly true in heavy elements where there are many core electrons that can contribute to the diamagnetism). Note however, if there are free spins in the material, then Curie paramagnetism occurs which is always stronger than any diamagnetism²⁷.

18.6.3 Curie Paramagnetism in Solids

Where to find free spins?

As discussed above, Curie paramagnetism describes the reorientation of free spins in an atom. We might ask how a “free spin” can occur in a solid? Our understanding of electrons in solids so far describes electrons as being either in full bands, in which case they cannot be flipped over at all; or in partially full bands, in which case the calculation of the Pauli susceptibility in section 4.3 is valid — albeit possibly with a modified density of states at the Fermi surface to reflect the details of the band structure (and with the Landau correction). So how is it that we can have a free spin?

Let us think back to the description of Mott insulators in section 15.4. In these materials, the Coulomb interaction between electrons is strong enough that no two electrons can double occupy the same site of the lattice. As a result, having one electron per site results in a “traffic jam” of electrons where no electron can hop to any other site. When this sort of Mott insulator forms, there is exactly one electron per site, which can be either spin-up or spin-down. Thus we have a free spin on each site exactly as we considered in the previous section!³¹

More generally we might expect that we could have some number N valence electrons per atom, which fill orbitals to form free spins as dictated by Hund’s rules. Again, if the Coulomb interaction is sufficiently strong that electrons cannot hop to neighboring sites, then the system will be Mott insulating and we can think of the spins as being free.

³¹This picture of a Mott insulator resulting in independent free spins will be examined more closely in chapter 22. Very weakly, some amount of (virtual) hopping can always occur and this will change the behavior at low enough temperatures.

Modifications of Free Spin Picture

Given that we have found free spins in a material, we can ask whether there substantial differences between a free spin in an isolated atom and a free spin in a material.

One possible modification is that the number of electrons on an atom become modified in a material. For example, we found above in section 18.2 that praseodymium (Pr) has three free electrons in its valence (4f) shell which form a total angular momentum of $J = 9/2$. However, in many compounds Pr exists as a +3 ion. In this case it turns out that both of the 6s electrons are donated as well as a single f electron. This leaves the Pr atom with two electrons in its f shell, thus resulting in a $J = 4$ angular momentum instead (you should be able to check this with Hund's rules).

Another possibility is that the atoms are no longer in a rotationally symmetric environment, they see the potential due to neighboring atoms, the so-called "crystal field". In this case orbital angular momentum is not conserved and the degeneracy of states all having the same L^2 is broken, a phenomenon known as *crystal field splitting*.

As a (very) cartoon picture of this physics, we can imagine a crystal which is highly tetragonal (see Fig. 11.11) where the lattice constant in one direction is quite different from the constant in the other two. We might imagine that an atom that is living inside such an elongated box would have a lower energy if its orbital angular momentum pointed along the long axes (say, the z-axis), rather than in some other direction. In this case, we might imagine that $L_z = +L$ and $L_z = -L$ might be lower energy than any of the other possible values of L .

Another thing that may happen due to crystal field splitting is that the orbital angular momentum may be pinned to have zero expectation (for example, if the ground state is a superposition of $L_z = +L$ and $L_z = -L$). In this case, the orbital angular momentum decouples from the problem completely (a phenomenon known as *quenching* of the orbital angular momentum), and the only magnetically active degrees of freedom are the spins. This is precisely what happens for most transition metals.³²

The most important moral to take home from this section is that paramagnets can have many different effective values of J , and one needs to know the microscopic details of the system before deciding which spin and orbital degrees of freedom are active.

18.7 Summary of Atomic Magnetism; Paramagnetism and Diamagnetism

- Susceptibility $\chi = dM/dH$ is positive for paramagnets and negative for diamagnets.
- Sources of paramagnetism: (a) Pauli paramagnetism of free electron gas (See section 4.3) (b) Free spin paramagnetism – know how to do the simple statmech exercise of calculating the paramagnetism of a free spin.
- The magnitude of the free spin determined by Hund's rules. The bonding of the atom, or environment of this atom (crystal field) can modify this result.

³²The 3d shell of transition metals is shielded from the environment only by the 4s electrons, whereas for rare earths the 4f shell is shielded by 6s and 5p. Thus the transition metals are much more sensitive to crystal field perturbations than the rare earths.

- Larmor diamagnetism can occur when atoms have $\mathbf{J} = \mathbf{0}$, therefore not having strong paramagnetism. This comes from the diamagnetic term of the Hamiltonian in first order perturbation theory. The diamagnetism per electron is proportional to the radius of the orbital.

References

- Ibach and Luth, section 8.1
- Hook and Hall, chapter 7
- Ashcroft and Mermin, chapter 31
- Kittel, chapter 11
- Blundell, chapter 2
- Burns, chapter 15A
- Goodstein, section 5.4a–c (doesn't cover diamagnetism)
- Rosenberg, chapter 11 (doesn't cover diamagnetism)

Chapter 19

Spontaneous Order: Antiferro-, Ferri-, and Ferro-Magnetism

In section 18.2.1 we commented that applying Hund’s rules to molecules can be quite dangerous since spins on neighboring atoms could favor either having their spins aligned or could favor having their spins anti-aligned depending on which of several effects is stronger. In chapter 22 below we will show models of how either behavior might occur. In this chapter we will assume there is an interaction between neighboring spins (a so-called *exchange* interaction, see footnote 18 from section 18.2.1) and we will explore how the interaction between neighboring spin aligns spins on a macroscopic scale.

We first assume that we have an insulator, i.e., electrons do not hop from site to site¹. We then write a model Hamiltonian as

$$\mathcal{H} = -\frac{1}{2} \sum_{i,j} J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j + \sum_i g\mu_B \mathbf{B} \cdot \mathbf{S}_i \quad (19.1)$$

where \mathbf{S}_i is the spin² on site i and \mathbf{B} is the magnetic field experienced by the spins³. Here $J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j$ is the interaction energy⁴ between spin i and spin j . Note that we have included a factor of 1/2 out front to avoid overcounting, since the sum actually counts both J_{ij} and J_{ji} (which are equal).

If $J_{ij} > 0$ then it is lower energy when spins i and j are aligned, whereas if $J_{ij} < 0$ then it is lower energy when the spins are anti-aligned.

The coupling between spins typically drops rapidly as the distance between the spins increases. A good model to use is one where only nearest neighbor spins interact with each other.

¹This might be the situation if we have a Mott insulator, as described in sections 15.4 and 18.6.3 above where strong interaction prevents electron hopping.

²When one discusses simplified models of magnetism, very frequently one writes angular momentum as \mathbf{S} without regards as to whether it is really \mathbf{S} , or \mathbf{L} or \mathbf{J} . It is also conventional to call this variable the “spin” even if it is actually from orbital angular momentum in a real material.

³Once again the plus sign in the final term assumes that we are talking about electronic moments. (See footnote 13 of section 4.3)

⁴WARNING: Many references use Heisenberg’s original convention that the interaction energy should be defined as $2J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j$ rather than $J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j$. However, more modern researchers use the latter, as we have here. This matches up with the convention used for the Ising model below Eq. 19.5, where the convention $2J$ is never used. At any rate, if someone on the street tells you J , you should ask whether they intend a factor of 2 or not.

Frequently one writes (neglecting the magnetic field \mathbf{B})

$$H = -\frac{1}{2} \sum_{i,j \text{ neighbors}} J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j$$

or using brackets $\langle i, j \rangle$ as a shorthand to indicate that i and j are neighbors,

$$H = -\frac{1}{2} \sum_{\langle i, j \rangle} J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j$$

In a uniform system where each spin is coupled to its neighbors with the same strength, we can drop the indices from J_{ij} (since they all have the same value) and we obtain the so-called *Heisenberg Hamiltonian*

$$H = -\frac{1}{2} \sum_{\langle i, j \rangle} J \mathbf{S}_i \cdot \mathbf{S}_j \quad (19.2)$$

19.1 (Spontaneous) Magnetic Order

As in the case of a ferromagnet, it is possible that even in the absence of any applied magnetic field, magnetism — or ordering of magnetic moments — may occur. This type of phenomenon is known as *spontaneous* magnetic order (since it occurs without application of any field). It is a subtle statistical mechanical question as to when magnetic interaction in a Hamiltonian actually results in spontaneous magnetic order. At our level of analysis we will assume that systems can always find ground states which “satisfy” the magnetic Hamiltonian. In chapter 21 we will consider how temperature might destroy this magnetic ordering.

19.1.1 Ferromagnets

As mentioned above, if $J > 0$ then neighboring spins want to be aligned. In this case the ground state is when all spins align together developing a macroscopic magnetic moment — this is what we call a *ferromagnet*, and is depicted on the left of Fig. 19.1. We will return to study these further in sections 20.1 and 21 below.

19.1.2 Antiferromagnets

On the other hand, if $J < 0$, neighboring spins want to point in opposite directions, and the most natural ordered arrangement is a periodic situation where alternating spins point in opposite directions, as shown on the right of Fig. 19.1 — this is known as an *antiferromagnet*. Such an antiferromagnet has zero net magnetization but yet is magnetically ordered. This type of antiperiodic ground state is sometimes known as a *Néel state* after Louis Néel who first proposed that these states exist⁵. We should be cautioned that our picture of spins pointing in directions is a classical picture, and is not quite right quantum mechanically. Particularly when the spin is small (like spin-1/2) the effects of quantum mechanics are strong and classical intuition can fail us. We will have a homework problem that shows that this classical picture of the antiferromagnet is not quite right, although it is fairly good when the spin on each site is larger than 1/2.

⁵Néel won a Nobel prize for this work in 1970.

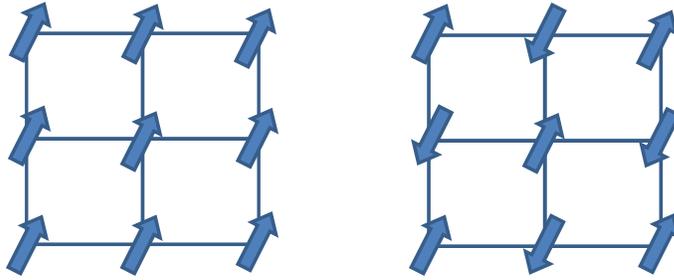


Figure 19.1: Magnetic Spin Orderings. Left: Ferromagnet — all spins aligned (at least over some macroscopic regions) giving finite magnetization. Middle: Antiferromagnet — Neighboring spins antialigned, but periodic. this so-called Néel state has zero net magnetization.

Detecting Antiferromagnetism with Diffraction

Being that antiferromagnets have zero net magnetization, how do we know they exist? What is their signature in the macroscopic world? For homework (see also section 21.2.2) we will explore a very nice method of determining that something is an antiferromagnet by examining its susceptibility as a function of temperature (in fact it was this type of experiment that Néel was analyzing when he realized that antiferromagnets exist). However, this method is somewhat indirect. A more direct approach is to examine the spin configuration using diffraction of neutrons. As mentioned in section 13.2, neutrons are sensitive to the spin direction of the object they scatter from. If we fix the spin polarization of an incoming neutron, it will scatter differently from the two different possible spin states of atoms in an antiferromagnet. The neutrons then see that the unit cell in this antiferromagnet is actually of size $2a$ where a is the distance between atoms (i.e., the distance between two atoms with the same spin is $2a$). Thus when the spins align antiferromagnetically, the neutrons will develop scattering peaks at reciprocal wavevectors $G = 2\pi/(2a)$ which would not exist if all the atoms were aligned the same way. This type of neutron diffraction experiment are definitive in showing that antiferromagnetic order exists⁶.

Frustrated Antiferromagnets

On certain lattices, for certain interactions, there is no ground state that fully “satisfies” the interaction for all spins. For example, on a triangular lattice if there is an antiferromagnetic interaction between bonds, there is no way that all the spins can point in the opposite direction from their neighbors. For example, as shown in the left of Fig. 19.2 on a triangle, once two of the spins are aligned opposite each other, independent of which direction the spin on the last site points, it cannot be antiparallel to both of its neighbors. It turns out that (assuming the spins are classical variables) the ground state of the antiferromagnetic Heisenberg Hamiltonian

⁶These are the experiments that won the Nobel prize for Clifford Schull. See footnote 9 from chapter 13.

on a triangle is the configuration shown on the right of Fig. 19.2. While each bond is not quite optimally anti-aligned, the overall energy is optimal for this Hamiltonian⁷

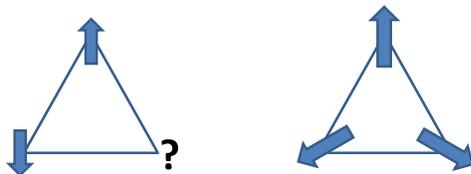


Figure 19.2: Cartoon of a Triangular Antiferromagnet. Left: An antiferromagnetic interaction on a triangular lattice is frustrated – not all spins can be antialigned with all of their neighbors. Right: The ground state of antiferromagnetic interaction on a triangle for classical spins (large S) is the state on the right, where spins are at 120° to their neighbor.

19.1.3 Ferrimagnetism

Once one starts to look for magnetic structure in materials, one can find many other interesting possibilities. One very common possibility is where you have a unit cell with more than one variety of atom, where the atoms have differing moments, and although the ordering is antiferromagnetic (neighboring spins point in opposite direction) there is still a net magnetic moment. An example of this is shown in Fig. 19.3. Here, the red atoms have a smaller moment than the green atoms and point opposite the green atoms. This type of configuration, where one has antiferromagnetic order, yet a net magnetization due to differing spin species, is known as *ferrimagnetism*. In fact, many of the most common magnets, such as magnetite (Fe_3O_4) are ferrimagnetic. Sometimes people speak of ferrimagnets as being a subset of ferromagnets (since they have nonzero net magnetic moment in zero field) whereas other people think the word “ferromagnet” implies that it exclude the category of ferrimagnets.⁸

19.2 Breaking Symmetry

In any of these ordered states, we have not yet addressed the question of which direction the spins will actually point. Strictly speaking, the Hamiltonian Eq. 19.2 is rotationally symmetric — the magnetization can point in any direction and the energy will be the same! In a real system, however, this is rarely the case: due to the asymmetric environment the atom feels within the lattice, there will be some directions that the spins would rather point than others (This physics

⁷Try showing this!

⁸The fact that the scientific cannot come to agreement on so many definitions does make life difficult sometimes. However, such disagreements inevitably come from the fact that many different communities, from high energy physicists to chemists, are interested in this type of physics. Coming from such diverse backgrounds, it is perhaps more surprising that there aren’t even more disagreements!

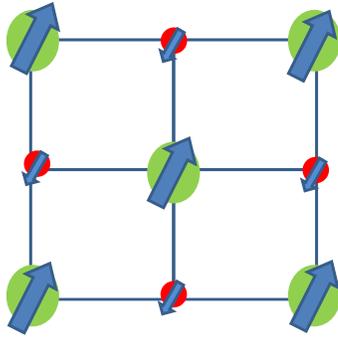


Figure 19.3: Cartoon of a Ferrimagnet: Ordering is antiferromagnetic, but because the different spin species have different moments, there is a net magnetization

was also discussed above in section 18.6). Thus to be more accurate we might need to add an additional term to the Heisenberg Hamiltonian. One possibility is to write⁹

$$H = -\frac{1}{2} \sum_{\langle i,j \rangle} J \mathbf{S}_i \cdot \mathbf{S}_j - \kappa \sum_i (S_i^z)^2 \quad (19.3)$$

(again dropping any external magnetic field). The κ term here favors the spin to be pointing in the $+\hat{z}$ direction or the $-\hat{z}$ direction, but not in any other direction. (You could imagine this being appropriate for a tetragonal crystal elongated in the \hat{z} direction). This energy from the κ term is sometimes known as the *anisotropy energy* since it favors certain directions over others. Another possible Hamiltonian is

$$H = -\frac{1}{2} \sum_{\langle i,j \rangle} J \mathbf{S}_i \cdot \mathbf{S}_j - \tilde{\kappa} \sum_i [(S_i^x)^4 + (S_i^y)^4 + (S_i^z)^4] \quad (19.4)$$

which favors the spin pointing along any of the orthogonal axis directions — but not towards any in-between angle.

In some cases (as we discussed in Section 18.6) the coefficient κ may be substantial. In other cases it may be very small. However, since the pure Heisenberg Hamiltonian Eq. 19.2 does not prefer any particular direction, even if the anisotropy (κ) term is extremely small, it will determine the direction of the magnetization in the ground state. We say that this term “breaks the symmetry.” Of course, there may be some symmetry remaining. For example, in Eq. 19.3, if the interaction is ferromagnetic, the ground state magnetization may be all spins pointing in the $+\hat{z}$ direction, or equally favorably, all spins pointing in the $-\hat{z}$ direction.

19.2.1 Ising Model

If the anisotropy (κ) term is extremely large, then this term can fundamentally change the Hamiltonian. For example, let us take a spin- S Heisenberg model. Adding the κ term in 19.3 with a

⁹For small values of the spin quantum number, these added interactions may be trivial. For example, for spin $1/2$, we have $(S_x)^2 = (S_y)^2 = (S_z)^2 = 1$. However, as S becomes larger, the spin becomes more like a classical vector and these such κ terms will favor the spin pointing in the corresponding directions.

large coefficient, forces the spin to be either $S_z = +S$ or $S_z = -S$ with all other values of S_z having a much larger energy. In this case, a new effective model may be written

$$\mathcal{H} = -\frac{1}{2} \sum_{\langle i,j \rangle} J \sigma_i \sigma_j + g \mu_B B \sum_i \sigma_i \quad (19.5)$$

where $\sigma_i = \pm S$ only (and we have re-introduced the magnetic field B). This model is known as the *Ising model*¹⁰ and is an extremely important model in statistical physics¹¹.

19.3 Summary of Magnetic Orders

- Ferromagnets: spins align. Antiferromagnets: spins antialign with neighbors so no net magnetization. Ferrimagnets: spins antialign with neighbors, *but* alternating spins are different magnitude so there is a net magnetization anyway. Microscopic spins structures of this sort can be observed with neutrons.
- Useful model Hamiltonians include Heisenberg ($-J \mathbf{S}_i \cdot \mathbf{S}_j$) for isotropic spins, and Ising $-J S_i^z S_j^z$ for spins that prefer to align along only one axis.
- Spins generally do not equally favor all directions equally (as the Heisenberg model suggest). Terms that favor spins along particular axes may be weak or strong. Even if they are weak, they will pick a direction among otherwise equally likely directions.

References

- Blundell, section 5.1–5.3 (Very nice discussion, but covers mean field theory at the same time which we will cover in chapter 21 below.)
- Burns, section 15.4–15.8 (same comment).

¹⁰“Ising” is properly pronounced “Ee-sing” or “Ee-zing”. In the United States it is habitually mispronounced “Eye-sing”. The Ising model was actually invented by Wilhelm Lenz (another example of Stigler’s law, see footnote 10 in section 14.2). Ising was the graduate student who worked on this model his graduate dissertation.

¹¹The Ising model is frequently referred to as the “hydrogen atom” of statistical mechanics since it is extremely simple, yet it shows many of the most important features of complex statistical mechanical systems. The one dimensional version of the model was solved in by Ising in 1925, and the two dimensional version of the model was solved by Onsager in 1944 (a chemistry Nobel Laureate, who was amusingly fired by my alma-mater, Brown University, in 1933). Onsager’s achievement was viewed as so important that Wolfgang Pauli wrote after world war two that “nothing much of interest has happened [in physics during the war] except for Onsagers exact solution of the two-dimensional Ising model.” (Perhaps Pauli was spoiled by the years of amazing progress in physics between the wars). If you are very brave, you might try calculating the free energy of the one-dimensional Ising model at finite temperature.

Chapter 20

Domains and Hysteresis

20.1 Macroscopic Effects in Ferromagnets: Domains

We might think that in a ferromagnet, all the spins in the system will align as described above in the Heisenberg (or Ising) models. However, in real magnets, this is frequently not the case. To understand why this is we imagine splitting our sample into two halves as shown in Fig. 20.1. Once we have two magnetic dipoles it is clear that they would be lower energy if one of them flipped over as shown it at the far right of Fig. 20.1. (The two north faces of these magnets repel each other¹). This energy, the long range dipolar force of a magnet, is not described in the Heisenberg or Ising model at all. In those models we have only included nearest neighbor interaction between spins. As we mentioned above, the actual magnetic dipolar force between electronic spins (or orbital moments) is tiny compared to the interaction “exchange” force between neighboring spins. But when you put together a whole lot of atoms (like 10^{23} of them!) to make a macroscopic magnet, the summed effect of their dipole moment can be substantial.

Of course, in an actual ferromagnet (or ferrimagnet), the material does not really break apart, but nonetheless different regions will have magnetization in different directions in order to minimize the dipolar energy. A region where the moments all point in one given direction is known as a *domain* or a *Weiss domain*.² The boundary of a domain, where the magnetization switches direction is known as a *domain wall*.³ Some possible examples of domain structures are sketched in Fig. 20.2. In the left two frames we imagine an Ising-like ferromagnet where the moment can only point up or down. The left most frame shows a magnet with net zero magnetization. Along the domain walls, the ferromagnetic Hamiltonian is “unsatisfied”. In other words, spin-up atoms on one side of the domain wall have spin-down neighbors — where the Hamiltonian says that they should want to have spin up neighbors only. What is happening is that the system is paying an

¹Another way to understand the dipolar force is to realize that the magnetic field far away from the magnets will be much lower if the two magnets are antialigned with each other. Since the electromagnetic field carries energy $\int dV |B|^2 / \mu_0$, minimizing this magnetic field lowers the energy of the two dipoles.

²After Pierre-Ernest Weiss, one of the fathers of the study of magnets from the early 1900s.

³Domain walls can also occur in antiferromagnets. Instead of the magnetization switching directions we imagine a situation where to the left of the wall, the up-spins are on the even sites, and the down-spins are on the odd-sites, whereas on the right of the domain wall the up-spins are on the odd sites and the down spins are on the even sites. At the domain wall, two neighboring sites will be aligned rather than anti-aligned. Since antiferromagnets have no net magnetization, the argument that domain walls should exist in ferromagnets is not valid for antiferromagnets. In fact, it is always energetically unfavorable for domain walls to exist in antiferromagnets, although they can occur at finite temperature.

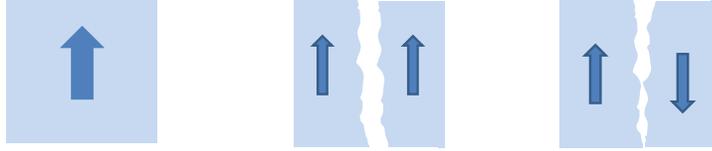


Figure 20.1: Dipolar Forces Create Magnetic Domains. Left: The original ferromagnet. Middle: The original ferromagnet broken into two halves. Right: Because two dipoles next to each other are lower energy if their moments are anti-aligned, the two broken halves would rather line up in opposing directions to lower their energies (the piece on the right hand side has been flipped over here). This suggests that in large ferromagnets, domains may form.

energy cost along the domain wall in order that the *global* energy associated with the long range dipolar forces is minimized.

If we apply a small up-pointing external field to this system, we will obtain the middle picture where the up pointing domains grow at the expense of the down pointing domains to give an overall magnetization of the sample. In the rightmost frame of Fig. 20.2 we imagine a sample where the moment can point along any of the crystal axis directions⁴. Again in this picture the total magnetization is zero but it has rather complicated domain structure.

20.1.1 Disorder and Domain Walls

The detailed geometry of domains in a ferromagnet depends on a number of factors. First of all, it depends on the overall geometry of the sample. (For example, if the sample is a very long thin rod and the system is magnetized along the long axis, it may gain very little energy by forming domains). It also depends on the relative energies of the neighbor interaction versus the long range dipolar interaction: increasing the strength of the long range dipolar forces with respect to the neighbor interaction will obviously decrease the size of domains (having no long range dipolar forces, will result in domains of infinite size). Finally, the detailed disorder in a sample can effect the shape and size of magnetic domains. For example, if the sample is polycrystalline, each domain could be a single crystallite (a single microscopic crystal).

⁴See for example the Hamiltonian, Eq. 19.4, which would have moments pointing only along the coordinate axes — although that particular Hamiltonian does not have the long range magnetic dipolar interaction written, so it would not form domains.

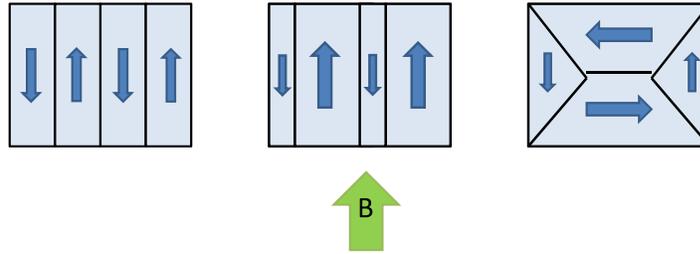


Figure 20.2: Some Possible Domain Structures for a Ferromagnet. Left: An Ising-like ferromagnet where in each domain the moment can only point either up or down. Middle: When an external magnetic field pointing upwards is applied to this ferromagnet, it will develop a net moment by having the down-domains shrink and the up-domains expand (The local moment per atom remains constant — only the size of the domains change). Right: In this ferromagnet, the moment can point in any of the crystal axis directions.

20.1.2 Disorder Pinning

Even for single-crystal samples, disorder can play an extremely important role in the physics of domains. For example, a domain-wall can have lower energy if it passes over a defect in the crystal. To see how this occurs let us look at a domain wall in an Ising ferromagnet as shown in Fig. 20.3. All bonds are marked red where spins are antialigned rather than aligned. In both figures the domain wall starts and ends at the same points, but on the right it follows a path through a defect in the crystal — in this case a site that is missing an atom. When it intersects the location of the missing atom, the number of antialigned bonds (marked) is lower, and therefore the energy is lower. Since this lower energy makes the domain wall stick to the missing site, we say the domain wall is *pinned* to the disorder.

20.1.3 The Bloch/Néel Wall

Our discussion of domain walls so far has assumed that the spins can only point up or down — that is, the κ term in Eq. 19.3 is extremely strong. However, it often happens that this is not true — the spins would prefer to point either up or down, but there is not a huge energy penalty for pointing in other directions instead. In this case the domain wall might instead be more of a smooth rotation from spins pointing up to spins pointing down as shown on the right of Fig. 20.4. This type of smooth domain wall is known as a *Bloch wall* or *Néel wall*⁵ depending on which direction the spin rotates in with respect to the direction of the domain wall itself (a somewhat subtle difference, which we will not discuss further here). The length of the domain wall (L in the

⁵We have already met our heroes of magnetism — Felix Bloch and Louis Néel.

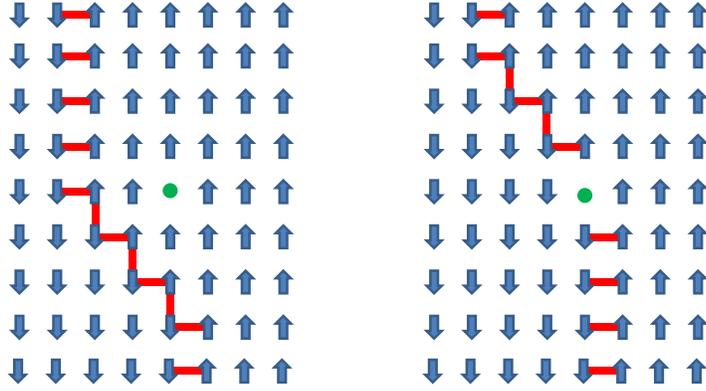


Figure 20.3: Domain Wall Pinning. The energy of a domain wall is lower if the domain wall goes through the position of a defect in the crystal. Here, the green dot is supposed to represent a missing spin. The red bonds, where spins are anti-aligned each cost energy. When the domain wall intersects the location of the missing spin, there are fewer red bonds, therefore it is a lower energy configuration. (There are 12 red bonds on the left, but only 10 on the right).

figure, i.e., how many spins are pointing neither up nor down) is clearly dependent on a balance between the J term of Eq. 19.3 (known sometimes as the *spin stiffness*) and the κ term, the anisotropy. As mentioned above, if κ/J is very large, then the spins must point either up or down only. In this case, the domain wall is very sharp, as depicted on the left of Fig. 20.4. On the other hand, if κ/J is small, then it costs little to point the spins in other directions, and it is more important that each spin points mostly in the direction of its neighbor. In this case, the domain wall will be very fat, as depicted on the right of Fig. 20.4.

A very simple scaling argument can give us an idea of how fat the Bloch/Néel wall is. Let us say that the length of the wall is N lattice constants, so $L = Na$ is the actual length of the twist in the domain wall (See Fig. 20.4). Roughly let us imagine that the spin twists uniformly over the course of these N spins, so between each spin and its neighbor, the spin twists an angle $\delta\theta = \pi/N$. The first term $-J\mathbf{S}_i \cdot \mathbf{S}_j$ in the Hamiltonian 19.3 then can be rewritten in terms of the angle between the neighbors

$$E_{one-bond} = -J\mathbf{S}_i \cdot \mathbf{S}_j = -JS^2 \cos(\theta_i - \theta_j) = -JS^2 \left(1 - \frac{(\delta\theta)^2}{2} + \dots \right)$$

where we have used the fact that $\delta\theta$ is small to expand the cosine. Naturally, the energy of this term is minimized if the two neighboring spins are aligned, that is $\delta\theta = 0$. However, if they are not quite aligned there is an energy penalty of

$$\delta E_{one-bond} = JS^2(\delta\theta)^2/2 = JS^2(\pi/N)^2/2$$

This is the energy per bond. So the energy of the domain wall due to this spin “stiffness” is

$$\frac{\delta E_{stiffness}}{A/a^2} = NJS^2(\pi/N)^2/2$$



Figure 20.4: Domain Wall Structure. Left: An infinitely sharp domain wall. This would be realized if the anisotropy energy (κ) is extremely large so the spins must point either up or down (i.e., this is a true Ising system). Right: A Bloch/Néel wall (actually this depicts a Néel wall) where the spin flips continuously from up to down over a length scale L . The anisotropy energy here is smaller so that the spin can point at intermediate angle for only a small energy penalty. By twisting slowly the domain wall will pay less spin-stiffness energy.

Here we have written the energy, per unit area A of the domain wall in units of the lattice constant a .

On the other hand, in Eq. 19.3 there is a penalty proportional to κS^2 per spin when the spins are not either precisely up or down. We estimate the energy due to this term to be κS^2 per spin, or a total of

$$\frac{\delta E_{anisotropy}}{A/a^2} \approx \kappa S^2 N$$

along the length of the twist.⁶ Thus the total energy of the domain wall is

$$\frac{E_{tot}}{A/a^2} = JS^2(\pi^2/2)/N + \kappa S^2 N$$

This can be trivially minimized resulting in a domain wall twist having length $L = Na$ with

$$N = C_1 \sqrt{(J/\kappa)} \quad (20.1)$$

⁶This approximation of the energy of the κ term is annoyingly crude. To be more precise, we should instead write $\kappa S^2 \cos^2(\theta_i)$ and then sum over i . Although this makes things more complicated, it is still possible to solve the problem so long as the spin twists slowly so that we can replace the finite difference $\delta\theta$ with a derivative, and replace the sum over sites with an integral. In this case, one minimizes the function

$$E = \int dx [JS^2(a^2/2)(d\theta(x)/dx)^2 - \kappa S^2 \cos^2 \theta(x)] / a$$

with a the lattice constant. Using calculus of variations the minimum of this energy is given by the solution of the differential equation

$$(Ja^2/\kappa)d^2\theta/dx^2 - \sin(2\theta) = 0$$

which has a truly remarkable solution of the form

$$\theta(x) = 2 \tan^{-1} \left(\exp \left[\sqrt{2}(x/a) \sqrt{\frac{\kappa}{J}} \right] \right)$$

where we once again see the same $L \sim \sqrt{J/\kappa}$ scaling. Plugging in this solution, the total energy of the domain wall becomes $E_{tot}/(A/a^2) = 2\sqrt{2}S^2\sqrt{J\kappa}$.

and a minimum domain wall energy per unit area

$$\frac{E_{tot}^{min}}{A/a^2} = C_2 S^2 \sqrt{J\kappa}$$

where C_1 and C_2 are constants of order unity (which we will not get right here considering the crudeness our approximation, but see footnote 6). As predicted, the length increases with J/κ . In many real materials the length of a domain wall can be hundreds of lattice constants.

Since the domain wall costs an energy per unit area, it is energetically unfavorable. However, as mentioned above, this energy cost is weighed against the long-range dipolar energy which tries to introduce domain walls. The more energy the domain wall costs, the larger individual domains will be (to minimize the number of domain walls). Note that if a crystal is extremely small (or, say, one considers a single crystallite within a polycrystalline sample) it can happen that the size of the crystal is much smaller than the optimum size of the domain wall twist. In this case the spins within this crystallite always stay aligned with each other.

Finally we comment that even though the actual domain wall may be hundreds of lattice constants thick, it is easy to see that these objects still have a tendency to stick to disorder as described in section 20.1.1 above.

20.2 Hysteresis in Ferromagnets

We know from our experience with electromagnetism that ferromagnets show a hysteresis loop with respect to the applied magnetic field, as shown in Fig. 20.5. After a large external magnetic field is applied, when the field is returned to zero there remains a residual magnetization. We can now ask why this should be true. In short, it is because there is a large activation energy for changing the magnetization.

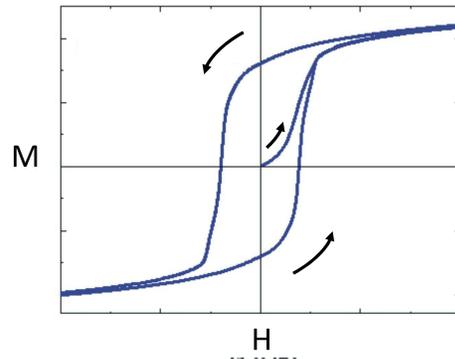


Figure 20.5: The Hysteresis Loop of a Ferromagnet

20.2.1 Single-Domain Crystallites

For example, let us consider the case of a ferromagnet made of many small crystallites. If the crystallites are small enough then all of the moments in each crystallite point in a single direction.

(We determined in section 20.1.3 that domain walls are unfavorable in small enough crystallites). So let us imagine that all of the microscopic moments (spins or orbital moments) in this crystallite are locked with each other and point in the same direction. The energy per volume of the crystallite in an external field can be written as

$$E/V = E_0 - \mathbf{M} \cdot \mathbf{B} - \kappa'(M^z)^2$$

where here \mathbf{M} is magnetization vector, and M_z is its component in the \hat{z} crystal axis. Here the anisotropy term κ' stems from the anisotropy term κ in the Hamiltonian 19.3⁷ Note that we have no J term since this would just give a constant if all the moments in the crystallite are always aligned with each other.

Assuming that the external field \mathbf{B} is pointing along the \hat{z} axis (although we will allow it to point either up or down) we then have

$$E/V = E_0 - |M||B| \cos \theta - \kappa'|M|^2 \cos^2 \theta \quad (20.2)$$

where $|M|$ is the magnitude of magnetization and $\cos \theta$ is the angle of the magnetization with respect to the \hat{z} axis.

We see that this energy is a parabola in the variable $\cos \theta$ which ranges from $+1$ to -1 . The minimum of this energy is always when the magnetization points in the same direction as the external field (which we have taken to always point in the either the $+\hat{z}$ or $-\hat{z}$ direction, corresponding to $\theta = 0$ or π). However, for small B_z the energy is not monotonic in θ . Indeed, having the magnetization point in the *opposite* direction as B is also a local minimum (because the κ' term favors pointing along the z -axis). This is shown schematically in Fig. 20.6. It is an easy exercise⁸ to show that there will be a local minimum of the energy with the magnetization pointing the opposite direction as the applied field for $B < B_{crit}$ with

$$B_{crit} = 2\kappa'|M|$$

So if the magnetization is aligned along the $-\hat{z}$ direction and a field $B < B_{crit}$ is applied in the $+\hat{z}$ direction, there is an activation barrier for the moments to flip over. Indeed, since the energy shown in Eq. 20.2 is an energy *per-volume*, the activation barrier⁹ can be very large. As a result, the moments will not be able to flip until a large enough field ($B > B_{crit}$) is applied to lower the activation barrier, at which point the moments flip over. Clearly this type of activation barrier can result in hysteretic behavior as shown in Fig. 20.5.

20.2.2 Domain Pinning and Hysteresis

Domains turn out to be extremely important for determining the detailed magnetic properties of materials – and in particular for understanding hysteresis in crystals that are sufficiently large that they are not single-domain (Recall that we calculated the size L of a domain wall in Eq. 20.1. Crystals larger than this size can in principle contain a domain wall). As mentioned above, when a magnetic field is externally applied to a ferromagnet, the domain walls move to re-establish a new domain configuration (See the left two panels of Fig. 20.2) and therefore a new magnetization.

⁷In particular since $\mathbf{M} = -g\mu_B\mathbf{S}\rho$ with ρ the number of spins per unit volume we have $\kappa' = \kappa/[(g\mu_B)^2\rho]$. Further we note that the $-\mathbf{M} \cdot \mathbf{B}$ term is precisely the Zeeman energy $+g\mu_B\mathbf{B} \cdot \mathbf{S}$ per unit volume.

⁸Try showing it!

⁹In principle the spins can get over the activation barrier either by being thermally activated or by quantum tunneling. However, if the activation barrier is sufficiently large (i.e., for a large crystallite) both of these are greatly suppressed.

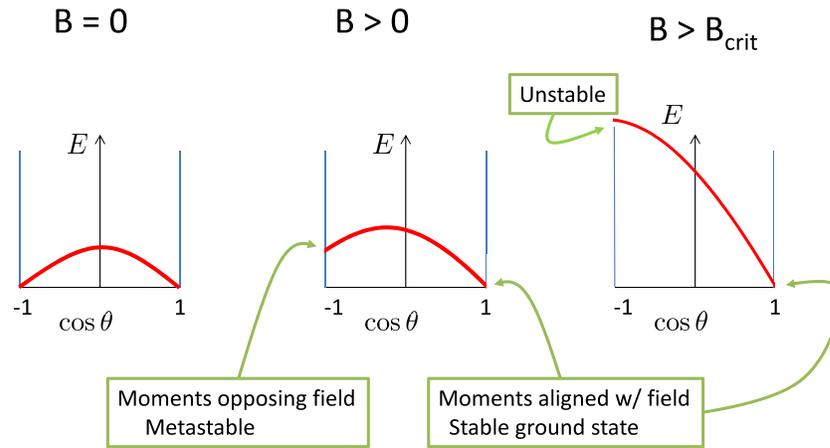


Figure 20.6: Energy of an Anisotropic Ferromagnet in a Magnetic Field as a Function of Angle. Left: Due to the anisotropy, in zero field the energy is lowest if the spins point either in the $+\hat{z}$ or $-\hat{z}$ direction. When a field is applied in the $+\hat{z}$ direction the energy is lowest when the moments are aligned with the field, but there is a metastable solution with the moments pointing in the opposite direction. The moments must cross an activation barrier to flip over. Right: For large enough field, there is no longer a metastable solution.

However, as we discussed in section 20.1.2 above, when there is disorder in a sample, the domain walls can get pinned to the disorder: There is a low energy configuration where the domain wall intersects the disorder, and there is then an activation energy to move the domain wall. This activation energy, analogous to what we found above in section 20.2.1, results in hysteresis of the magnet.

It is frequently the case that one wants to construct a ferromagnet which retains its magnetization extremely well — i.e., where there is strong hysteresis, and even in the absence of applied magnetic field there will be a large magnetization. This is known as a “hard” magnet (also known as a “permanent” magnet). It turns out that much of the trick of constructing hard magnets is arranging to insert appropriate disorder and microstructure to strongly pin the domain walls.

20.3 Summary of Domains and Hysteresis in Ferromagnets

- Although short range interaction in ferromagnet favors all magnetic moments to align, long range dipolar forces favors spins to anti-align. A compromise is reached with domains of aligned spins where different domains point in different directions. A very small crystal may be a single domain.
- The actual domain wall boundary may be a continuous rotation of the spin rather than a

sudden flip over a single bond-length. The size of this spin structure depends on the ratio of the ferromagnetic energy to the anisotropy energy. (I.e., if it is very costly to have spins point in directions between up and down then the wall will be over a single bond length).

- Domain walls are lower energy if they intersect certain types of disorder in the solid. This results in the *pinning* of domain walls – they stick to the disorder.
- In a large crystal, changes in magnetization occur by changing the size of domains. In polycrystalline samples with very small crystallites, changes in magnetization occur by flipping over individual single-domain crystallites. Both of these processes can require an activation energy (domain motion requires activation energy if domain walls are pinned) and thus result in hysteretic behavior a magnetization in ferromagnets.

References

- Hook and Hall, section 8.7
- Blundell, section 6.7
- Burns, section 15.10
- Ashcroft and Mermin, end of chapter 33

Also good (but covers material in random order compared to what we want):

- Rosenberg, chapter 12
- Kittel, chapter 12

Chapter 21

Mean Field Theory

Given a Hamiltonian for a magnetic system, we are left with the theoretical task of how to predict its magnetization as a function of temperature (and possibly external magnetic field). Certainly at low temperature, the spins will be maximally ordered, and at high temperature, the spins will thermally fluctuate and will be disordered. But calculating the magnetization as a function of temperature and applied magnetic field, is typically a very hard task. Except for a few very simple exactly solvable models (like the Ising model in one dimension) we must always resort to approximations. The most important and probably the simplest such approximation is known as “Mean Field Theory” or “Molecular Field theory” or “Weiss Mean Field theory”¹ which we will discuss in depth in this chapter.

The general concept of mean field theory proceeds in two steps:

- First, one examines one site (or one unit cell, or some small region) and treats it exactly. Any object outside the unit cell is approximated as an expectation (an average or a *mean*).
- The second step is to impose self-consistency: Every site (or unit cell, or small region) in the entire system should look the same. So the one site we treated exactly should have the same average as all of the others.

This procedure is extremely general and can be applied to problems ranging from magnetism to liquid crystal to fluid mechanics. We will demonstrate the procedure as it applies to ferromagnetism. For a homework problem we will consider how mean field theory can be applied to antiferromagnets as well (further generalizations should then be obvious).

21.1 Mean Field Equations for the Ferromagnetic Ising Model

As an example, let us consider the spin-1/2 Ising model

$$\mathcal{H} = -\frac{1}{2} \sum_{\langle i,j \rangle} J \sigma_i \sigma_j + g \mu_B B \sum_i \sigma_j$$

¹The same Pierre-Ernest Weiss for whom Weiss domains are named.

where $J > 0$, and here $\sigma = \pm 1/2$ is the z -component of the spin and the magnetic field B is applied in the \hat{z} direction (and as usual μ_B is the Bohr magneton). For a macroscopic system, this is a statistical mechanical system with 10^{23} degrees of freedom, where all the degrees of freedom are now coupled to each other. In other words, it looks like a hard problem!

To implement mean field theory, we focus in on one site of the problem, say, site i . The Hamiltonian for this site can be written as

$$\mathcal{H}_i = \left(g\mu_B B - J \sum_j \sigma_j \right) \sigma_i$$

where the sum is over sites j that neighbor i . We think of the term in brackets as being caused by some effective magnetic field seen by the spin on site i , thus we define $B_{eff,i}$ such that

$$g\mu_B B_{eff,i} = g\mu_B B - J \sum_j \sigma_j$$

with again j neighboring i . Now $B_{eff,i}$ is not a constant, but is rather an operator since it contains the variables σ_j which can take several values. However, the first principle of mean field theory is that we should simply take an average of all quantities that are not site i . Thus we write the Hamiltonian of site i as

$$\mathcal{H}_i = g\mu_B \langle B_{eff} \rangle \sigma_i$$

This is precisely the same Hamiltonian we considered when we studied paramagnetism in Eq. 18.6 above, and it is easily solvable. In short, one writes the partition function

$$Z_i = e^{-\beta g\mu_B \langle B_{eff} \rangle / 2} + e^{\beta g\mu_B \langle B_{eff} \rangle / 2}$$

From this we can derive the expectation of the spin on site i (compare Eq. 18.8)

$$\langle \sigma_i \rangle = -\frac{1}{2} \tanh(\beta g\mu_B \langle B_{eff} \rangle / 2) \quad (21.1)$$

However, we can also write that

$$g\mu_B \langle B_{eff} \rangle = g\mu_B B - J \sum_j \langle \sigma_j \rangle$$

The second step of the mean field approach is to set $\langle \sigma \rangle$ to be equal on all sites of the lattice, so we obtain

$$g\mu_B \langle B_{eff} \rangle = g\mu_B B - Jz \langle \sigma \rangle \quad (21.2)$$

where z is the number of neighbors j of site i (this is known as the *coordination number* of the lattice, and this factor has replaced the sum on j). Further, again assuming that $\langle \sigma \rangle$ is the same on all lattice sites, from Eq. 21.1 and 21.2, we obtain the self-consistency equation for $\langle \sigma \rangle$ given by

$$\langle \sigma \rangle = -\frac{1}{2} \tanh(\beta [g\mu_B B - Jz \langle \sigma \rangle] / 2) \quad (21.3)$$

The expected moment per site will correspondingly be given by².

$$m = -g\mu_B \langle \sigma \rangle \quad (21.4)$$

²Recall the the spin points opposite the moment! Ben Franklin, why do you torture us so? (See footnote 13 of section 4.3)

21.2 Solution of Self-Consistency Equation

The self-consistency equation, Eq. 21.3 is still complicated to solve. One approach is to find the solution graphically. For simplicity, let us set the external magnetic field B to zero. We then have the self-consistency equation

$$\langle \sigma \rangle = \frac{1}{2} \tanh \left(\frac{\beta J z}{2} \langle \sigma \rangle \right) \quad (21.5)$$

We then choose a value of the parameter $\beta J z / 2$. Let us start by choosing a value $\beta J z / 2 = 1$ that is somewhat small, i.e., a high temperature. Then in Fig. 21.1 we plot both the right hand side of Eq. 21.5 as a function of $\langle \sigma \rangle$ (in blue) and the left hand side of Eq. 21.5 (in green). Note that the left hand side is $\langle \sigma \rangle$ so the straight line is simply $y = x$. We see that there is only a single point where the two curves meet, i.e., where the left side equals the right side. This point, in this case is $\langle \sigma \rangle = 0$. From this we conclude that, for this value temperature, within mean field approximation, there is no magnetization in zero field.

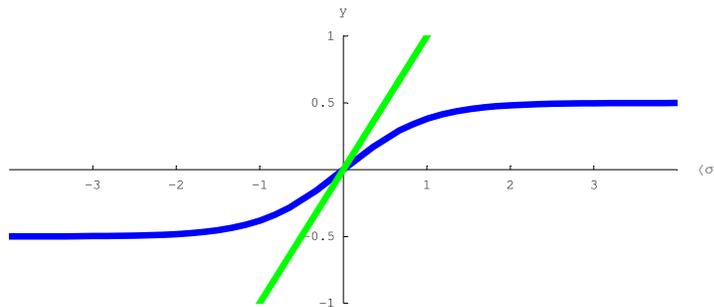


Figure 21.1: Graphical Solution of the Mean Field Self Consistency Equations at Relatively High Temperature $\beta J z / 2 = 1$. The blue line is the tanh of Eq. 21.5. The green line is just the line $y = x$. Eq. 21.5 is satisfied only where the two curves cross – i.e., at $\langle \sigma \rangle = 0$ meaning that at this temperature, within the mean field approximation, there is no magnetization.

Let us now reduce the temperature substantially to $\beta J z / 2 = 6$. Analogously, in Fig. 21.2 we plot both the right hand side of Eq. 21.5 as a function of $\langle \sigma \rangle$ (in blue) and the left hand side of Eq. 21.5 (in green). Here, however, we see there are three possible self-consistent solutions to the equations. There is the solution at $\langle \sigma \rangle = 0$ as well as two solutions marked with arrows in the figure at $\langle \sigma \rangle \approx \pm 0.497$. The two nonzero solutions tell us that at low temperature this system can have nonzero magnetization even in the absence of applied field — i.e., it is ferromagnetic.

The fact that we have possible solutions with the magnetization pointing in both directions is quite natural: The Ising ferromagnet can be polarized either spin up or spin down. However, the fact that there is also a self-consistent solution with zero magnetization at the same temperature seems a bit puzzling. We will see as a homework assignment that when there are three solutions,

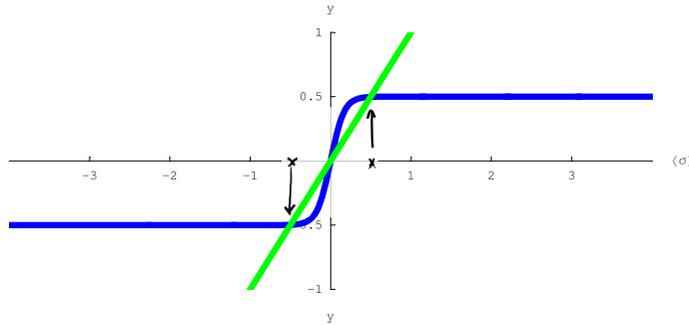


Figure 21.2: Graphical Solution of the Mean Field Self Consistency Equations at Relatively Low Temperature $\beta Jz/2 = 6$. Here, the curves cross at three possible values ($\langle \sigma \rangle = 0$ and $\langle \sigma \rangle \approx \pm .497$). The fact that there is a solution of the self-consistency equations with nonzero magnetization tells us that the system is ferromagnetic (the zero magnetization solution is non-physical).

the zero magnetization solution is actually a solution of maximal free energy not minimal free energy, and therefore should be discarded³.

Thus the picture that arises is that at high temperature the system has zero magnetization (and we will see below that it is paramagnetic) whereas at low temperature a nonzero magnetization develops and the system becomes ferromagnetic⁴. The transition between these two behaviors occurs at a temperature known as T_c , which stands for *critical temperature*⁵ or *Curie temperature*⁶. It is clear from Figs. 21.1 and 21.2 that the behavior changes from one solution to three solutions precisely when the straight green line is tangent to the tanh curve, i.e., when the slope of the tanh is unity. This tangency condition thus determines the critical temperature. Expanding the tanh for small argument, we obtain the tangency condition

$$1 = \frac{1}{2} \left(\frac{\beta_c Jz}{2} \right)$$

or when the temperature is

$$k_b T_c = \frac{Jz}{4}$$

Using the above technique, one can solve the self-consistency equations (Eq. 21.5) at any temperature (although there is no nice analytic expression, it can be solved numerically or

³We will see (as a homework problem) that our self-consistency equations are analogous to when we find the minimum of a function by differentiation — and we may also find maxima as well.

⁴It is quite typical that at high temperatures, a ferromagnet will turn into a paramagnet, unless something else happens first — like the crystal melts.

⁵Strictly speaking it should only be called a critical temperature if the transition is second order. I.e., if the magnetization turns on continuously at this transition. For the Ising model, this is in fact true, but for some magnetic systems it is not true.

⁶Named for Pierre again.

graphically). The results are shown in Fig. 21.3. Note that at low enough temperature, all of the spins are fully aligned ($\langle\sigma\rangle = 1/2$ which is the maximum possible for a spin-1/2). One can also, in

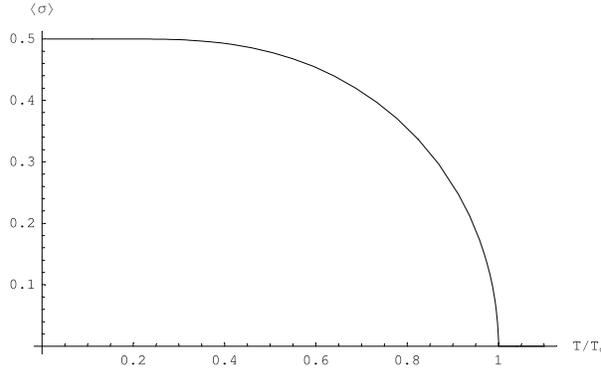


Figure 21.3: Magnetization as a Function of Temperature. The plot shows the magnitude of the moment per site in units of $g\mu_B$ as a function of temperature in the mean field approximation of the spin-1/2 Ising model, with zero external magnetic field applied.

principle, solve the self-consistency equation (Eq. 21.3) with finite magnetic field \mathbf{B} .

21.2.1 Paramagnetic Susceptibility

At high temperature there will be zero magnetization in zero externally applied field. However, at finite field, we will have a finite magnetization. Let us imagine applying a small magnetic field and solve the self-consistency equations Eq. 21.3. Since the applied field is small, we can assume that the induced $\langle\sigma\rangle$ is also small. Thus we can expand the tanh in Eq.21.3 to obtain

$$\langle\sigma\rangle = \frac{1}{2}(\beta[Jz\langle\sigma\rangle - g\mu_B B]/2)$$

Rearranging this then gives

$$\langle\sigma\rangle = -\frac{\frac{1}{4}(\beta g\mu_B)B}{1 - \frac{1}{4}\beta Jz} = -\frac{\frac{1}{4}(g\mu_B)B}{k_b(T - T_c)}$$

which is valid only so long as $\langle\sigma\rangle$ remains small. The moment per site is then given by (See Eq. 21.4) $m = -g\mu_B\langle\sigma\rangle$ which divided by the volume of a unit cell gives the magnetization M . Thus we find that the susceptibility is

$$\chi = \mu_0 \frac{\partial M}{\partial B} = \frac{\frac{1}{4}\rho(g\mu_B)^2\mu_0}{k_b(T - T_c)} = \frac{\chi_{Curie}}{1 - T_c/T} \quad (21.6)$$

where ρ is the number of spins per unit volume and χ_{Curie} is the pure Curie susceptibility of a system of (noninteracting) spin-1/2 particles (Compare Eq. 18.9). Eq. 21.6 is known as the *Curie-Weiss Law*. Thus, we see that a ferromagnet above its critical temperature is roughly a paramagnet with an enhanced susceptibility. Note that the susceptibility diverges at the transition temperature when the system becomes ferromagnetic.⁷

⁷This divergence is in fact physical. As the temperature is reduced towards T_c , the divergence tells us that it takes a smaller and smaller applied B field to create some fixed magnetization M . This actually makes sense since

21.2.2 Further Thoughts

As mentioned above, the mean-field procedure is actually very general. As a homework problem we will also study the antiferromagnet. In this case, we divide the system into two sublattices — representing the two sites in a unit cell. In that example we will want to treat one spin of each sublattice exactly, but as above each spin sees only the average field from its neighbors. One can generalize even further to consider very complicated unit cells.

Aside: It is worth noting that the result of solving the Antiferromagnetic Ising model gives

$$\chi = \frac{\chi_{Curie}}{1 + T_c/T}$$

compared Eq. 21.6. It is this difference in susceptibility that pointed the way to the discovery of antiferromagnets.

We see that in both the ferromagnetic and antiferromagnetic case, at temperatures much larger than the critical temperature (much larger than the exchange energy scale J), the system behaves like a pure free spin Curie paramagnet. In section 18.6.3 above we asked where we might find free spins so that a Curie paramagnet might be realized. In fact, now we discover that any ferromagnet or antiferromagnet (or ferrimagnet for that matter) will appear to be free spins at temperatures high enough compared to the exchange energy. Indeed, it is almost always the case that when one thinks that one is observing free spins, at low enough energy scales one discovers that in fact the spins are coupled to each other!

The principle of mean field theory is quite general and can be applied to a vast variety of difficult problems in physics⁸. No matter what the problem, the principle remains the same — isolate some small part of the system to treat exactly, average everything outside of that small system, then demand self-consistency: that the average of the small system should look like the chosen average of the rest of the system.

While the mean field approach is merely an approximation, it is frequently a very good approximation for capturing a variety of physical phenomena. Furthermore, many of its shortcomings can be systematically improved by considering successively more corrections to the initial mean field approach⁹.

21.3 Summary of Mean Field Theory

- Understand the mean field theory calculation for ferromagnets. Understand how you would generalize this to any model of antiferromagnets (homework), ferrimagnets (try it!) different spins, anisotropic models, etc etc.
- For the ferromagnet the important results of mean field theory includes:

once the temperature is below T_c , the magnetization will be nonzero even in the absence of any applied B .

⁸In chapter 2 we already saw another example of mean field theory, when we considered the Boltzmann and Einstein models of specific heat of solids. There we considered each atom to be in a harmonic well formed by all of its neighbors. The single atom was treated exactly, whereas the neighboring atoms were treated only approximately in that their positions were essentially averaged in order to simply form the potential well — and nothing further was said of the neighbors. Another example in similar spirit was given in footnote 2 of chapter 17 where an alloy of Al and Ga with As is replaced by some averaged atom $\text{Al}_x\text{Ga}_{1-x}$ and is still considered a periodic crystal. (No unit cell is treated exactly in this case, all are replaced by the average unit cell).

⁹The motivated student might want to think about various ways one might improve mean field theory systematically. One approach is discussed in the Additional Problems.

- a finite temperature phase transition from a low temperature ferromagnetic phase to a high temperature paramagnetic phase at a transition temperature known as the Curie temperature.
- Above the curie temperature the paramagnetic susceptibility is $\chi = \chi_{Curie}/(1 - T_c/T)$ where χ_0 is the susceptibility of the corresponding model where the ferromagnetic coupling between sites is turned off.
- Below T_c the magnetic moment turns on, and increases to saturation at the lowest temperature.

References on Mean Field Theory

- Ibach and Luth, chapter 8 (particularly 8.6, 8.7)
- Hook and Hall, chapter 8 (particularly 8.3, 8.4)
- Kittel, beginning of chapter 12
- Burns, section 15.5
- Ashcroft and Mermin, chapter 33

Chapter 22

Magnetism from Interactions: The Hubbard Model

So far we have only discussed ferromagnetism in the context of isolated spins on a lattice that align due to their interactions with each other. However, in fact many materials have magnetism where the magnetic moments – the aligned spins – are not pinned down, but rather can wander through the system. This phenomenon is known as *itinerant ferromagnetism*¹. For example, it is easy to imagine a free electron gas where the number of up spins is different from the number of down spins. However, for completely free electrons it is always lower energy to have the same number of up and down spins than to have the numbers differ². So how does it happen that electrons can decide, even in absence of external magnetic field, to polarize their spins? The culprit is the strong Coulomb interaction between electrons. On the other hand, we will see that antiferromagnetism can also be caused by strong interaction between electrons as well!

The Hubbard model³ is an attempt to understand the magnetism that arises from interactions between electrons. It is certainly the most important model of interacting electrons in modern condensed matter physics. We will see through this model how interactions can produce both ferro- and anti-ferromagnetism (this was alluded to in section 18.2.1).

The model is relatively simple to describe⁴. First we write a tight binding model for a band

¹Itinerant means traveling from place to place without a home (from Latin *iter*, or *itiner* meaning journey or road. In case anyone cares.)

²The total energy of having N electrons spin up in a system is proportional to $NE_F = N(N/V)^{2/d}$ where d is the dimensionality of the system (you should be able to prove this easily). We can write $E = CN^{1+a}$ with $a > 0$ and C some constant. For N_\uparrow up spins and N_\downarrow downspins, we have a total energy $E = CN_\uparrow^{1+a} + CN_\downarrow^{1+a} = C(N_\uparrow^{1+a} + (N - N_\uparrow)^{1+a})$ where N is the total number of electrons. Setting $dE/dN = 0$ immediately gives $N_\uparrow = N/2$ as the minimum energy configuration.

³John Hubbard, a British physicist, wrote down this model in 1963 and it quickly became an extremely important example in the attempt to understand interacting electrons. Despite the success of the model, Hubbard, who died relatively young in 1980, did not live to see how important his model became: In 1986, when the phenomenon of “high temperature superconductivity” was discovered by Bednorz and Müller (resulting in a Nobel prize the following year), the community quickly came to believe that an understanding of this phenomenon would only come from studying the Hubbard model. Over the next two decades the Hubbard model took on the status of being *the* most important question in condensed matter physics. Its complete solution remains elusive despite the tens of thousands of papers written on the subject. It is a shame that we do not have time to discuss superconductivity in this course.

⁴The reason most introductory books do not cover the Hubbard model is that the model is conventionally introduced using so-called “second quantized” notation — that is, using field-theoretic methods which are rather

of electrons as we did in chapter 10 with hopping parameter t . (We can choose to do this in one, two, or three dimensions as we see fit⁵). We will call this Hamiltonian H_0 . As we derived above (and should be easy to derive in two and three dimensions now) the full bandwidth of the band is $4dt$ in d dimensions. We can add as many electrons as we like to this band. Let us define the number of electrons in the band per site to be called the doping, x (so that $x/2$ is the fraction of k states in the band which are filled being that there are two spin states). As long as we do not fill all of the states in the band ($x < 2$), in the absence of interactions, this partially filled tight binding band is a metal. Finally we include the Hubbard interaction

$$H_{interaction} = \sum_i U n_{i\uparrow} n_{i\downarrow} \quad (22.1)$$

where here $n_{i\uparrow}$ is the number of electrons with spin up on site i and $n_{i\downarrow}$ is the number of electrons on site i with spin down, and $U > 0$ is an energy known as the repulsive Hubbard interaction energy. This term gives an energy penalty of U whenever two electrons sit on the same site of the lattice. This short ranged interaction term is an approximate representation of the Coulomb interaction between electrons. The full Hubbard model Hamiltonian is given by the sum of the kinetic and interaction pieces

$$H = H_0 + H_{interaction}$$

22.1 Ferromagnetism in the Hubbard Model

Why should this on-site interaction create magnetism? Imagine for a moment that all of the electrons in the system had the same spin state (a so-called “spin-polarized” configuration). If this were true, by the Pauli exclusion principle, no two electrons could ever sit on the same site. In this case, the expectation of the Hubbard interaction term would be zero

$$\langle \text{Polarized Spins} | H_{interaction} | \text{Polarized Spins} \rangle = 0$$

which is the lowest possible energy that this interaction term could have. On the other hand, if we filled the band with only one spin-species, then the Fermi energy (and hence the kinetic energy of the system) would be much higher than if the electrons could be distributed between the two possible spin states. Thus, it appears that there will be some competition between the potential and kinetic energy that decides whether the spins align or not.

22.1.1 Hubbard Ferromagnetism Mean Field Theory

To try to decide quantitatively whether spins will align or not we start by writing

$$U n_{i\uparrow} n_{i\downarrow} = \frac{U}{4} (n_{i\uparrow} + n_{i\downarrow})^2 - \frac{U}{4} (n_{i\uparrow} - n_{i\downarrow})^2$$

Now we make the approximation of treating all operators $n_{i,\uparrow}$ and $n_{i,\downarrow}$ as their expectations.

$$U n_{i\uparrow} n_{i\downarrow} \approx \frac{U}{4} \langle n_{i\uparrow} + n_{i\downarrow} \rangle^2 - \frac{U}{4} \langle n_{i\uparrow} - n_{i\downarrow} \rangle^2$$

advanced. We will avoid this approach, but as a result, we cannot delve too deep into the physics of the model.

⁵In one dimension, the Hubbard model is exactly solvable.

This type of approximation is a type of mean-field theory, similar to that we encountered in the previous chapter 21: We replace operators by their expectations.⁶ The expectation $\langle n_{i\uparrow} + n_{i\downarrow} \rangle$ in the first term is just the average number of electrons on site i which is just the average number of particles per site,⁷ which is equal to the doping x , which we keep fixed.

Correspondingly, the second expectation, $\langle n_{i\uparrow} - n_{i\downarrow} \rangle$, is related to the magnetization of the system. In particular, since each electron carries⁸ a magnetic moment of μ_B , the magnetization⁹ is

$$M = (\mu_B/v)\langle n_{i\downarrow} - n_{i\uparrow} \rangle$$

with v the volume of the unit cell. We thus see that the expectation of the energy of the Hubbard interaction is given by

$$\langle H_{interaction} \rangle \approx (V/v)(U/4) (x^2 - (Mv/\mu_B)^2) \quad (22.2)$$

where V/v is the number of sites in the system. Thus, as expected, increasing the magnetization M decreases the expectation of the interaction energy. To determine if the spins actually polarize we need to weigh this potential energy gain against the kinetic energy cost.

22.1.2 Stoner Criterion¹⁰

Here we calculate the kinetic energy cost of polarizing the spins in our model and we balance this against the potential energy gain. We will recognize this calculation as being almost identical to the calculation we did way back in section 4.3 when we studied Pauli paramagnetism (but we repeat it here for clarity).

Consider a system (at zero temperature for simplicity) with the same number of spin up and spin down electrons. Let $g(E_F)$ be the total density of states at the Fermi surface per unit volume (for both spins put together). Now, let us flip over a small number of spins so that the spin up and spin down Fermi surfaces have slightly different energies¹¹.

$$\begin{aligned} E_{F,\uparrow} &= E_F + \delta\epsilon/2 \\ E_{F,\downarrow} &= E_F - \delta\epsilon/2 \end{aligned}$$

The difference in the number density of up and down electrons is then

$$\rho_{\uparrow} - \rho_{\downarrow} = \int_0^{E_F + \delta\epsilon/2} dE \frac{g(E)}{2} - \int_0^{E_F - \delta\epsilon/2} dE \frac{g(E)}{2}$$

where we have used the fact that the density of states per unit volume for either the spin-up or spin-down species is $\frac{g(E)}{2}$.

⁶This is a slightly different type of mean field theory from that encountered in chapter 21. Previously we considered some local degree of freedom (some local spin) which we treated exactly, and replaced all other spins by their average. Here, we are going to treat the kinetic energy term exactly, but replace the operators in the potential energy term by their averages.

⁷This assumes that the system remains homogeneous — that is, that all sites have the same average number of electrons.

⁸We have assumed an electron g-factor of $g = 2$ and an electron spin of $1/2$. Everywhere else in this chapter the symbol g will only be used for density of states.

⁹Recall magnetization is moment per unit volume.

¹⁰This has nothing to do with the length of your dreadlocks or the number of Grateful Dead shows you have been to (I've been to 6 shows ... I think).

¹¹If we were being very careful we would adjust E_F to keep the overall electron density $\rho_{\uparrow} + \rho_{\downarrow}$ fixed as we change $\delta\epsilon$. For small $\delta\epsilon$ we would find that E_F remains unchanged as we change $\delta\epsilon$ but this is not true for larger $\delta\epsilon$.

Although we could carry forward at this point and try to perform the integrals generally for arbitrary $\delta\epsilon$ (indeed we will have a homework problem on this) it is enough for our present discussion to consider the simpler case of very small $\delta\epsilon$. In this case, we have

$$\rho_{\uparrow} - \rho_{\downarrow} = \delta\epsilon \frac{g(E_F)}{2}$$

The difference in the number of up and down electrons is related to the magnetization of the system by⁸

$$M = \mu_B(\rho_{\downarrow} - \rho_{\uparrow})$$

so

$$M = -\mu_B \delta\epsilon \frac{g(E_F)}{2}$$

The kinetic energy per unit volume is a bit more tricky. We write

$$\begin{aligned} K &= \int_0^{E_F + \delta\epsilon/2} dE E \frac{g(E)}{2} + \int_0^{E_F - \delta\epsilon/2} dE E \frac{g(E)}{2} \\ &= 2 \int_0^{E_F} dE E \frac{g(E)}{2} + \int_{E_F}^{E_F + \delta\epsilon/2} dE E \frac{g(E)}{2} - \int_{E_F - \delta\epsilon/2}^{E_F} dE E \frac{g(E)}{2} \end{aligned} \quad (22.3)$$

$$\begin{aligned} &\approx K_{M=0} + \frac{g(E_F)}{2} \left[\left(\frac{(E_F + \delta\epsilon/2)^2}{2} - \frac{E_F^2}{2} \right) - \left(\frac{E_F^2}{2} - \frac{(E_F - \delta\epsilon/2)^2}{2} \right) \right] \\ &= K_{M=0} + \frac{g(E_F)}{2} (\delta\epsilon/2)^2 \\ &= K_{M=0} + \frac{g(E_F)}{2} \left(\frac{M}{\mu_B g(E_F)} \right)^2 \end{aligned} \quad (22.4)$$

Where $K_{M=0}$ is the kinetic energy per unit volume for a system with no net magnetization (equal numbers of spin-up and spin-down electrons).

We can now add this result to Eq. 22.2 to give the total energy of the system per unit volume

$$E_{tot} = E_{M=0} + \left(\frac{M}{\mu_B} \right)^2 \left[\frac{1}{2g(E_F)} - \frac{vU}{4} \right]$$

with v the volume of the unit cell. We thus see that for

$$U > \frac{2}{g(E_F)v}$$

the energy of the system is lowered by increasing the magnetization from zero. This condition for itinerant ferromagnetism is known as the *Stoner criterion*¹².

Aside: We did a lot of work to arrive at Eq. 22.4. In fact, we could have almost have written it down with no work at all based on the calculation of the Pauli susceptibility we did back in section 4.3. Recall first that when an external magnetic field is applied in the up direction to a system, there is an energy induced from the coupling of the spins to the field which is given by $\mu_B(\rho_{\uparrow} - \rho_{\downarrow})B = -MB$ (with positive M being defined

¹²Edmund Stoner was a British physicist who, among other things, figured out the Pauli Exclusion principle in 1924 a year before Pauli. However, Stoner's work focused on the spectra, and behavior, of atoms, and he was not bold enough to declare the exclusion was a fundamental property of electrons. Stoner was diagnosed with diabetes in 1919 at 20 years of age and grew progressively weaker for the next eight years. In 1927 Insulin treatment became available, saving his life. He died in 1969.

in the same direction as positive B so that having the two aligned is low energy). Also recall in section 4.3 that we derived the (Pauli) susceptibility of an electron system is

$$\chi_{Pauli} = \mu_0 \mu_B^2 g(E_F)$$

which means that when a magnetic field B is applied, a magnetization $\chi_{Pauli} B / \mu_0$ is induced. Thus we can immediately conclude that the energy of such a system in an external field *must* be of the form

$$E(M) = \frac{M^2 \mu_0}{2\chi_{Pauli}} - MB$$

To see that this is correct, we minimize the energy with respect to M at a given B and we discover that this properly gives us $M = \chi_{Pauli} B / \mu_0$. Thus, at zero applied B , the energy should be

$$E(M) = \frac{M^2 \mu_0}{2\chi_{Pauli}} = \frac{M^2}{2\mu_B^2 g(E_F)}$$

exactly as we found in Eq. 22.4!

22.2 Mott Antiferromagnetism in the Hubbard Model

In fact, the Hubbard model is far more complex than the above mean field calculation would lead one to believe. Let us now consider the case where the doping is such that there is exactly one electron per site of the lattice. For noninteracting electrons, this would be a half-filled band, and hence a conductor. However, if we turn on the Hubbard interaction with a large U , the system becomes an insulator. To see this, imagine one electron sitting on every site. In order for an electron to move, it must hop to a neighboring site which is already occupied. This process therefore costs energy U , and if U is large enough, the hopping cannot happen. This is precisely the physics of the Mott insulator which we discussed above in section 15.4.

With one immobile electron on each site we can now ask which way the spins align (in the absence of external field). For a square or cubic lattice, there are two obvious options: either the spins want to be aligned with their neighbors or they want to be anti-aligned with their neighbors (ferromagnetism or antiferromagnetism). It turns out that antiferromagnetism is favored! To see this, consider the antiferromagnetic state $|GS_0\rangle$ shown on the left of Fig. 22.1. In the absence of hopping this state is an eigenstate with zero energy (as is any other state where there is precisely one electron on each site). We then consider adding the hopping perturbatively. Because the hopping Hamiltonian allows an electron to hop from site to site (with hopping amplitude $-t$), the electron can make a “virtual” hop to a neighboring site, as shown in the right of Fig. 22.1. The state on the right $|X\rangle$ is of higher energy, (in the absence of hopping it has energy U because of the double occupancy). Using second order perturbation theory we obtain

$$\begin{aligned} E(|GS_0\rangle + \text{hopping}) &= E(|GS_0\rangle) + \sum_X \frac{|\langle X|H_{hop}|GS_0\rangle|^2}{E_{GS_0} - E_X} \\ &= E(|GS_0\rangle) - \frac{Nz|t|^2}{U} \end{aligned}$$

In the first line the sum is over all $|X\rangle$ states that can be reached in a single hop from the state $|GS_0\rangle$. In the second line, we have counted the number of such terms to be Nz where z is the coordination number (number of nearest neighbors) and N is the number of sites. Further we have inserted $-t$ for the amplitude of hopping from one site to the next. Note that if the spins were all aligned, no virtual intermediate state $|X\rangle$ could exist since it would violate the Pauli exclusion

principle (hopping of electrons conserves spin state, so spins cannot flip over during a hop, so there is strictly no double occupancy). Thus we conclude that the antiferromagnetic state has its energy lowered compared to the ferromagnetic state in the limit of large U in a Mott insulating phase.

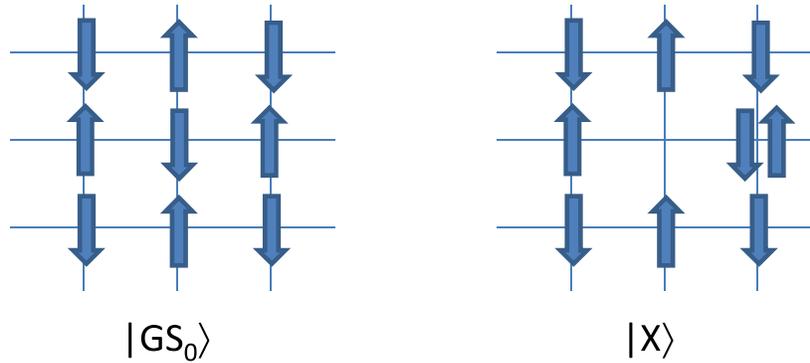


Figure 22.1: Spin Configurations of the Half Filled Hubbard Model. Left: The proposed antiferromagnetic ground state in the limit that t is very small. Right: A higher energy state in the limit of small t which can occur by an electron from one site hopping onto a neighboring site. The energy penalty for double occupancy is U .

Admittedly the above argument appears a bit handwaving (It is correct though!). To make the argument more precise, one should be much more careful about how one represents states with multiple electrons. This typically requires field theoretic techniques. A very simple example of how this is done (without more advanced techniques) is presented in the appendix to this chapter.

Nonetheless, the general physics of why the antiferromagnetic Mott insulator state should be lower energy than its ferromagnetic counterpart can be understood qualitatively without resorting to the more precise arguments. On each site one can think of an electron as being confined by the interaction with its neighbors to that site. In the ferromagnetic case, the electron cannot make any excursions to neighboring sites because of the Pauli exclusion principle (these states are occupied). However, in the antiferromagnetic case, the electron can make excursions, and even though when the electron wanders onto neighboring sites, the energy is higher, there is nonetheless some amplitude for this to happen.¹³ Allowing the electron wavefunction to spread out always lowers its energy¹⁴.

Indeed, in general a Mott insulator (on a square or cubic lattice) is typically an antiferromagnet (unless some other interesting physics overwhelms this tendency). It is generally believed that there is a substantial range of t , U and doping x where the ground state is antiferromagnetic. Indeed, many real materials are thought to be examples of antiferromagnetic Mott insulators. In-

¹³Similar to when a particle is in a potential well $V(x)$, there is some amplitude to find the electron at a position such that $V(x)$ is very large.

¹⁴By increasing Δx we can decrease Δp and thus lower the kinetic energy of the particle, as per the Heisenberg uncertainty principle.

terestingly, it turns out that in the limit of very very strong on-site interaction $U \rightarrow \infty$ adding even a single additional hole to the half-filled Mott insulator will turn the Mott antiferromagnet into a ferromagnet! This rather surprising result due to Nagaoka and Thouless¹⁵ (one of the few key results about the Hubbard model which is known as a rigorous theorem) shows the general complexity of this model.

22.3 Summary of the Hubbard Model

- Hubbard model includes tight-binding hopping t and on-site “Hubbard” interaction U
- For partially filled band, the repulsive interaction (if strong enough) makes the system an (itinerant) ferromagnet: aligned spins can have lower energy because they do not double occupy sites, and therefore are lower energy with respect to U although it costs higher kinetic energy to align all the spins.
- For a half-filled band, the repulsive interaction makes the Mott insulator antiferromagnetic: virtual hopping lowers the energy of anti-aligned neighboring spins.

References on Hubbard Model

Unfortunately there are essentially no references that I know of that are readable without background in field theory and second quantization.

22.4 Appendix: The Hubbard model for the Hydrogen Molecule

Since my above perturbative calculation showing antiferromagnetism is very hand-waving, I thought it useful to do a real (but very simple) calculation showing how, in principle, these calculations are done more properly. This appendix is certainly nonexaminable, but if you are confused about the above discussion of antiferromagnetism in the Hubbard model, this appendix might be enlightening to read.

The calculation given here will address the Hubbard model for the Hydrogen molecule. Here we consider two nuclei A and B near each other, with a total of two electrons – and we consider only the lowest spatial orbital (the s-orbital) for each atom¹⁶. There are then four possible states which an electron can be in:

$$A \uparrow \quad A \downarrow \quad B \uparrow \quad B \downarrow$$

To indicate that we have put electron 1 in, say the $A \uparrow$ state, we write the wavefunction

$$|A \uparrow\rangle \longleftrightarrow \varphi_{A\uparrow}(1)$$

(Here φ is the wavefunction (1) is shorthand for the position \mathbf{r}_1 as well as the spin σ_1 coordinate).

¹⁵David Thouless, born in Scotland, is one of the most prominent names in modern condensed matter physics. He has not yet won a Nobel prize, but he is frequently mentioned as a high contender. Yosuki Nagaoka is a prominent Japanese theorist.

¹⁶This technique can in principle be used for any number of electrons in any number of orbitals although exact solution becomes difficult as the Schroedinger matrix becomes very high dimensional and hard to diagonalize exactly necessitating sophisticated approximation methods.

For a two electron state, we are only allowed to write wavefunctions that are overall antisymmetric. So given two single electron orbitals α and β (α and β take values in the four possible orbitals $A \uparrow, A \downarrow, B \uparrow, B \downarrow$) we write so-called Slater determinants to describe the antisymmetric two particle wavefunctions

$$|\alpha; \beta\rangle = \frac{1}{\sqrt{2}} \det \begin{vmatrix} \alpha(1) & \beta(1) \\ \alpha(2) & \beta(2) \end{vmatrix} = (\alpha(1)\beta(2) - \beta(1)\alpha(2))/\sqrt{2} = -|\beta; \alpha\rangle$$

Note that this Slater determinant can be generalized to write a fully antisymmetric wavefunction for any number of electrons. If the two orbitals are the same, then the wavefunction vanishes (as it must by Pauli exclusion).

For our proposed model of the Hydrogen molecule, we thus have six possible states for the two electrons

$$\begin{aligned} |A \uparrow; A \downarrow\rangle &= -|A \downarrow; A \uparrow\rangle \\ |A \uparrow; B \uparrow\rangle &= -|B \uparrow; A \uparrow\rangle \\ |A \uparrow; B \downarrow\rangle &= -|B \downarrow; A \uparrow\rangle \\ |A \downarrow; B \uparrow\rangle &= -|B \uparrow; A \downarrow\rangle \\ |A \downarrow; B \downarrow\rangle &= -|B \downarrow; A \downarrow\rangle \\ |B \uparrow; B \downarrow\rangle &= -|B \downarrow; B \uparrow\rangle \end{aligned}$$

The Hubbard interaction energy (Eq. 22.1) is diagonal in this basis — it simply gives an energy penalty U when there are two electrons on the same site. We thus have

$$\langle A \uparrow; A \downarrow | H_{interaction} | A \uparrow; A \downarrow \rangle = \langle B \uparrow; B \downarrow | H_{interaction} | B \uparrow; B \downarrow \rangle = U$$

and all other matrix elements are zero.

To evaluate the hopping term we refer back to where we introduced tight binding in section 5.3.2 and chapter 10. Analogous to that discussion, we see that the hopping term with amplitude $-t$ turns an $A \uparrow$ orbital into a $B \uparrow$ orbital or vice versa, and similarly turns a $A \downarrow$ into a $B \downarrow$ and vice versa (The hopping does not change the spin). Thus, for example, we have

$$\langle A \downarrow; B \uparrow | H_{hop} | A \downarrow; A \uparrow \rangle = -t$$

where here the hopping term turned the B into an A . Note that this implies similarly that

$$\langle A \downarrow; B \uparrow | H_{hop} | A \uparrow; A \downarrow \rangle = t$$

since $|A \downarrow; A \uparrow\rangle = -|A \uparrow; A \downarrow\rangle$.

Since there are six possible basis states, our most general Hamiltonian can be expressed as a six by six matrix. We thus write our Schroedinger equation as

$$\begin{pmatrix} U & 0 & -t & t & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -t & 0 & 0 & 0 & 0 & -t \\ t & 0 & 0 & 0 & 0 & t \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -t & t & 0 & U \end{pmatrix} \begin{pmatrix} \psi_{A \uparrow A \downarrow} \\ \psi_{A \uparrow B \uparrow} \\ \psi_{A \uparrow B \downarrow} \\ \psi_{A \downarrow B \uparrow} \\ \psi_{A \downarrow B \downarrow} \\ \psi_{B \uparrow B \downarrow} \end{pmatrix} = E \begin{pmatrix} \psi_{A \uparrow A \downarrow} \\ \psi_{A \uparrow B \uparrow} \\ \psi_{A \uparrow B \downarrow} \\ \psi_{A \downarrow B \uparrow} \\ \psi_{A \downarrow B \downarrow} \\ \psi_{B \uparrow B \downarrow} \end{pmatrix}$$

where here we mean that the full wavefunction is the sum

$$\begin{aligned} |\Psi\rangle &= \psi_{A \uparrow A \downarrow} |A \uparrow; A \downarrow\rangle + \psi_{A \uparrow B \uparrow} |A \uparrow; B \uparrow\rangle + \psi_{A \uparrow B \downarrow} |A \uparrow; B \downarrow\rangle \\ &+ \psi_{A \downarrow B \uparrow} |A \downarrow; B \uparrow\rangle + \psi_{A \downarrow B \downarrow} |A \downarrow; B \downarrow\rangle + \psi_{B \uparrow B \downarrow} |B \uparrow; B \downarrow\rangle \end{aligned}$$

We note immediately that the Hamiltonian is block diagonal. We have eigenstates

$$|A \uparrow; B \uparrow\rangle \quad |A \downarrow; B \downarrow\rangle$$

both with energy $E = 0$ (hopping is not allowed and there is no double occupancy, so no Hubbard interaction either). The remaining four by four Shroedinger equation is then

$$\begin{pmatrix} U & t & -t & 0 \\ t & 0 & 0 & t \\ -t & 0 & 0 & -t \\ 0 & t & -t & U \end{pmatrix} \begin{pmatrix} \psi_{A\uparrow A\downarrow} \\ \psi_{A\uparrow B\downarrow} \\ \psi_{A\downarrow B\uparrow} \\ \psi_{B\uparrow B\downarrow} \end{pmatrix} = E \begin{pmatrix} \psi_{A\uparrow A\downarrow} \\ \psi_{A\uparrow B\downarrow} \\ \psi_{A\downarrow B\uparrow} \\ \psi_{B\uparrow; B\downarrow} \end{pmatrix}$$

We find one more eigenvector $\propto (0, 1, 1, 0)$ with energy $E = 0$ corresponding to the state¹⁷

$$\frac{1}{\sqrt{2}} (|A \uparrow; B \downarrow\rangle + |A \downarrow; B \uparrow\rangle)$$

A second eigenstate has energy U and has a wavefunction

$$\frac{1}{\sqrt{2}} (|A \uparrow; A \downarrow\rangle - |B \uparrow; B \downarrow\rangle)$$

The remaining two eigenstates are more complicated, and have energies $\frac{1}{2} (U \pm \sqrt{U^2 + 16t^2})$. The ground state, always has energy

$$E_{ground} = \frac{1}{2} (U - \sqrt{U^2 + 16t^2})$$

In the limit of t/U becoming zero, the ground state wavefunction becomes very close to

$$\frac{1}{\sqrt{2}} (|A \uparrow; B \downarrow\rangle - |A \downarrow; B \uparrow\rangle) + \mathcal{O}(t/U) \quad (22.5)$$

with amplitudes of order t/U for the two electrons to be on the same site. In this limit the energy goes to

$$E_{ground} = -4t^2/U$$

which is almost in agreement with our above perturbative calculation — the prefactor differs from that mentioned in the above calculation by a factor of 2. The reason for this discrepancy is that the ground state is not just \uparrow on one site and \downarrow on the other, but rather a superposition between the two. This superposition can be thought of as a (covalent) chemical bond (containing two electrons) between the two atoms.

In the opposite limit, $U/t \rightarrow 0$ the ground state wavefunction for a single electron is the symmetric superposition $(|A\rangle + |B\rangle)/\sqrt{2}$ (see section 5.3.2) assuming $t > 0$. This is the so-called “bonding” orbital. So the ground state for two electrons is just the filling of this bonding orbital with both spins — resulting in

$$\begin{aligned} \frac{|A \uparrow\rangle + |B \uparrow\rangle}{\sqrt{2}} \otimes \frac{|A \downarrow\rangle + |B \downarrow\rangle}{\sqrt{2}} &= \frac{1}{2} (|A \uparrow; A \downarrow\rangle + |A \uparrow; B \downarrow\rangle + |B \uparrow; A \downarrow\rangle + |B \uparrow; B \downarrow\rangle) \\ &= \frac{1}{2} (|A \uparrow; A \downarrow\rangle + |A \uparrow; B \downarrow\rangle - |A \downarrow; B \uparrow\rangle + |B \uparrow; B \downarrow\rangle) \end{aligned}$$

Note that eliminating the double occupancy states (simply crossing them out)¹⁸ yields precisely the same result as Eq. 22.5. Thus, as the interaction is turned on it simply suppresses the double occupancy in this case.

¹⁷The three states with $E = 0$ are in fact the $S_z = -1, 0, 1$ states of $S = 1$. Since the Hamiltonian is rotationally invariant, these all have the same energy.

¹⁸Eliminating doubly occupied orbitals by hand is known as *Gutzwiller projection* (after Martin Gutzwiller) and is an extremely powerful approximation tool for strongly interacting systems.

Chapter 23

Magnetic Devices

This is the chapter on magnetic devices. It is *NONEXAMINABLE* It is also *NONFINISHED*. I hope to finish this soon!

Indices

These notes have two indices¹.

In the index of people, Nobel laureates are marked with *. There are over 50 of them. A few stray celeb pop stars got into the index as well.

A few people whose names are mentioned did not end up in the index because the use of their name is so common that it is not worth indexing them as people as well. A few examples are Coulomb's law; Fourier transform; Boltzmann's constant; Taylor expansion; Hamiltonian; Jacobian, and so forth. But then again, I did index Schroedinger Equation and Fermi Statistics under Schroedinger and Fermi respectively. So I'm not completely consistent. So sue me.

The index of topics was much more difficult to put together. It was hard to figure out what the most useful division of topics is to put in the index. I tried to do this so that the index would be maximally useful - but I'm not sure how good a job I did. Most book indices are not very useful, and now I know why — it is hard to predict why a reader is going to want to look something up.

¹Making it a tensor. har har.

Index of People

- Anderson, Philip*, 1, 2
 Appleton, Edward*, 150
- Bardeen, John**, 46
 Bednorz, Johannes*, 235
 Berg, Moe, 141
 Bethe, Hans*, 27
 Bloch, Felix*, 36, 160–161, 219–222
 Bohr, Niels*, 33, 132, 181, 195, 200, 228
 Boltzmann, Ludwig, 8, 17, 19, 20, 67, 184, 232
 Born, Max*, 12–13, 48
 Bose, Satyendra, 9, 33, 72, 75
 Bragg, William Henry*, 132–135, 140–141, 143–144, 147
 Bragg, William Lawrence*, 132–135, 140–141, 143–144, 147
 Braun, Karl Ferdinand*, 191
 Bravais, Auguste, 99, 107, 110–111, 123
 Brillouin, Leon, 69, 71, 74, 75, 80–86, 95, 123–127, 153–161, 163, 165–170, 172
 Brockhouse, Bertram*, 136, 147
- Crick, Francis*, 147
 Curie, Marie**, 46, 204
 Curie, Pierre*, 204, 205, 230–232
- Darwin, Charles Galton, 143
 Darwin, Charles Robert, 143
 de Hevesy, George*, *see* Hevesy, George*
 Debye, Peter*, 11–17, 24, 27, 41, 67, 71, 73–74, 80, 141, 143
 Deisenhofer, Johann*, 147
 Dirac, Paul*, 27–30, 33, 132, 175
 Drude, Paul, 19–26, 32, 35, 37, 178–179, 184
 Dulong, Pierre, 7–8, 10, 15
- Earnshaw, Samuel, 196
 Ehrenfest, Paul, 19
 Einstein, Albert*, 8–11, 13, 17, 33, 67, 73, 207, 232
 Very Smart, 11
- Faraday, Michael, 196
 Fawcett, Farrah, 48
 Fermi, Enrico*, 24, 27–30, 132–133, 184
 Floquet, Gaston, 160
- Frankenheim, Moritz, 111
 Franklin, Benjamin, 33, 203, 228
 Franklin, Rosalind, 147
 Franz, Rudolph, 24, 25, 32, 37
 Fuller, Richard Buckminster, 58
- Galton, Francis, 143
 Geim, Andre*, 196
 Gell-Mann, Murray*, ii
 Gutzwiller, Martin, 243
- Hall, Edwin, 21–23, 179
 Heath, James*, 58
 Heisenberg, Werner*, 27, 33, 211, 212, 215, 217, 240
 Hevesy, George*, 132
 Higgs, Peter, 2
 Hodgkin, Dorothy*, 147
 Hubbard, John, 235–243
 Huber, Robert*, 147
 Hund, Friedrich, 197–205, 211
- Ising, Ernst, 211, 215–217, 219, 227–232
- Kendrew, John*, 147
 Kepler, Johannes, 107
 Klechkovsky, Vsevolod, 198
 Kohn, Walter*, 44
 Kronecker, Leopold, 116
 Kroto, Harold*, 58
- Landau, Lev*, 33, 37, 170, 205, 208
 Langevin, Paul, 204
 Larmor, Joseph, 205–208
 Laue, Max von*, 132–135, 140–141
 Laughlin, Robert*, 3
 Leeuwen, Hendrika van, 195
 Lenz, Heinrich, 196
 Lenz, Wilhelm*, 216
 Lipscomb, William*, 147
 Lorentz, Hendrik*, 20, 24, 32, 143, 207
 Lorenz, Ludvig, 24
- Madelung, Erwin, 45, 198
 Magnés, Shephard, 195
 Marconi, Guglielmo*, 191
 Mather, John*, 12
 Merton, Robert, 160

- Michel, Harmut*, 147
Miller, William Hallows, 119–123
Mott, Nevill*, 170, 208, 211, 239, 240
Müller, Karl Alex*, 235
Mulliken, Robert*, 46, 197
- Néel, Louis*, 212–213, 219–222
Nagaoka, Yosuki, 241
Newton, Isaac, 33, 60, 68, 177, 178
Newton-John, Irene Born, 48
Newton-John, Olivia, 48
Noether, Emmy, 75
- Onsager, Lars*, 216
Oppenheimer, J. Robert, 48
- Pauli, Wolfgang*, 24, 27, 32–34, 37, 45, 178, 201, 205, 207–208, 216, 236, 238, 240, 242
Pauling, Linus**, 27, 46
Peierls, Rudolf, 27
Peltier, Jean Charles, 24–26
Perutz, Max*, 147
Petit, Alexis, 7–8, 10, 15
Planck, Max*, 12–14, 17
Poisson, Siméon, 117
Pople, John*, 44, 49
- Rabi, Isadore Isaac*, 27
Riemann, Bernhard, 14
Rutherford, Ernest Lord*, 150
Rydberg, Johannes, 181
- Sanger, Fredrick**, 46
Scherrer, Paul, 141
Schroedinger, Erwin*, 3, 11, 33, 41–42, 44, 48–51, 88–90, 242
Schull, Clifford*, 136, 147, 213
Seebeck, Thomas, 25
Seitz, Fredrick, 103–104, 109, 110, 112
Simon, Steven H., 1
Slater, John, 28, 242
Smalley, Richard*, 58
Smoot, George*, 12
Sommerfeld, Arnold, 26–37, 41
Spears, Britney, iii
Stigler, Stephen, 160, 216
Stoner, Edmund, 237–239
Stormer, Horst*, 3
Superfluid, 207
- Thomson, Joseph John*, 19
Thouless, David, 241
Travolta, John, 48
Tsui, Dan*, 3
- Van der Waals, J. D.*, 53–54
van Leeuwen, Hendrika, *see* Leeuwen, Hendrika van
Van Vleck, John*, 205, 206
Von Karman, Theodore, 12–13
von Laue, Max*, *see* Laue, Max von*
- Waller, Ivar, 143
Watson, James*, 147
Weiss, Pierre, 217, 227, 231
Wiedemann, Gustav, 24, 25, 32, 37
Wigner, Eugene*, 33, 103–104, 109, 110, 112
Wilson, Kenneth*, 2
- Zeeman, Pieter*, 33, 203, 223

Index of Topics

- Acceptor, 179, 182, 183
- Acoustic Mode, 81, 86
- Adiabatic Demagnetization, 206
- Alloy, 189
- Amorphous Solid, 59, 147
- Anderson-Higgs Mechanism, 2
- Anisotropy Energy, 215, 219–221
- Antibonding Orbital, 47, 51–53
- Antiferromagnetism, 212–213, 216, 232
 - Frustrated, 213–214
 - Mott, *see* Mott Antiferromagnetism, 241
- Atomic Form Factor, *see* Form Factor
- Aufbau Principle, 197

- Band, *see* Band Structure
- Band Gap, 94, 96, 158, 161, 163, 174
 - Designing of, 189–190
 - Direct, *see* Direct Gap
 - Indirect, *see* Indirect Gap
 - Non-Homogeneous, 190
- Band Insulator, 95, 96, 163, 167, 168, 174, 179
- Band Structure, 90–96, 158, 161, 163–170
 - Engineering, 189
 - Failures of, 170
 - of Diamond, 125
- Bandwidth, 90
- Basis
 - in Crystal Sense, 79, 85, 104–106, 112
 - Vectors
 - Primitive, *see* Primitive Lattice Vectors
- BCC Lattice, *see* Body Centered Cubic Lattice
- Bloch Function, 160
- Bloch Wall, 219–222
- Bloch's Theorem, 36, 160–161
- Body Centered Cubic Lattice, 107–109, 112, 141
 - Miller Indices, 120
 - Selection Rules, 138–139
- Bohr Magneton, 33, 200, 203, 204, 228
- Boltzmann Model of Solids, 8, 17, 67, 232
- Boltzmann Statistics, 184, 188
- Boltzmann Transport Equation, 20
- Bonding Orbital, 47, 51–53, 243
- Books
 - Good, iii–iv
- Born-Oppenheimer Approximation, 48
- Born-Von-Karman Boundary Condition, *see* Periodic Boundary Conditions
- Bose Occupation Factor, 9, 72, 75
- Bragg Condition, 132–135, 140–141, 143–144, 149
- Bravais Lattice
 - Nomenclatural Disagreements, 99
- Bravais Lattice Types, 110–111
- Brillouin Zone, 69, 71, 74, 75, 80–86, 96, 123–127, 153–161, 163, 167, 170, 172
 - Boundary, 71, 84, 95, 96, 155–161, 163, 166, 167
 - Definition of, 69, 124
 - First, 69, 71, 74, 75, 84, 123–125, 161, 167–169
 - Definition of, 124
 - Number of k States in, 124
 - Second, 84, 124, 167, 168
 - Definition of, 124
- Buckyball, 58
- Bulk Modulus, 65

- Carrier Freeze Out, 182, 188
- Chemical Bond, 41–56
 - Covalent, *see* Covalent Bond
 - Fluctuating Dipolar, *see* Van der Waals Bond
 - Hydrogen, *see* Hydrogen Bond
 - Ionic, *see* Ionic Bond
 - Metallic, *see* Metallic Bond
 - Molecular, *see* Van der Waals Bond
 - Van der Waals, *see* Van der Waals Bond
- Compressibility, 64, 82
- Condensed Matter
 - Definition of, 1
- Conduction Band, 163, 164, 173, 175, 179
- Conductivity
 - of Metals, 21
 - Thermal, *see* Thermal Conductivity
- Conventional Unit Cell, 102, 108, 110, 112
 - For BCC Lattice, 108
 - of FCC Lattice, 110
- Coordination Number, 109
- Cornstarch, 60
- Covalent Bond, 42–44, 46–52, 243

- Critical Temperature, 230
- Crystal Field, 209
- Crystal Momentum, 74–75, 92, 95, 132, 153, 161, 172
- Crystal Plane, *see* Lattice Plane
- Cubic Lattice, *see* Simple Cubic or FCC or BCC
- Curie Law, 204, 205, 231
- Curie Temperature, 230
- Curie-Weiss Law, 231
- Curse, 173

- Debye Frequency, 13, 14
- Debye Model of Solids, 11–17, 41, 67, 71–74, 80–143
- Debye Temperature, 14
- Debye-Scherrer Method, *see* Powder Diffraction
- Debye-Waller Factor, 143
- Density of States
 - Electronic, 31, 33, 183, 184, 237
 - of Debye Model, 13
 - of One Dimensional Vibration Model, 73
- Diamagnetism, 209
 - Definition of, 196
 - Landau, 33, 205, 208
 - Larmor, 205–208, 210
- Diffraction, 133–134, 213
- Dipole Moment, *see* Electric Dipole Moment or Magnetic Dipole Moment
- Dirac Equation, 175
- Direct Band Gap, 173
- Direct Gap, 164, 171–172, 175
- Direct Lattice, 70
- Direct Transition, 171–173
- Dispersion Relation
 - of Electrons, 90
 - of Vibrational Normal Modes, 68
 - of Vibrations, 81
- DNA, 55, 59, 147
- Dollars
 - One Million, 14
- Domain Wall, 217–222, 225
- Domains, 217–225
- Donor, 179, 182, 183
- Doped Semiconductor, 179–182, 186–187
- Doping, *see* Impurities
- Doughnut Universe, 12
- Drude Model of Electron Transport, 19–27, 35, 37, 178–179, 185, 188
 - Shortcomings of, 25
- Dulong-Petit Law, 7, 8, 10, 15, 17

- Earnshaw's Theorem, 196
- Effective Mass, 91, 96, 158, 175–177
- Effective mass, 187
- Einstein Frequency, 8, 10
- Einstein Model of Solids, 8–11, 15, 17, 67, 73–74, 232
- Einstein Temperature, 10
- Elasticity, 64
- Electric Dipole Moment, 53
- Electric Susceptibility
 - See* Polarizability, 53
- Electron
 - g-factor, *see* g-factor of Electron
- Electron Affinity, 44–46
 - Table of, 45
- Electron Donor, *see* Donor
- Electron Mobility, 179
- Electron Transport
 - Drude Model, *see* Drude Model of Electron Transport
- Electronegativity, 42, 46
 - Mulliken, 46
- Energy Band, *see* Band Structure
- Exchange Interaction, 202, 211, 217
- Extended Zone Scheme, 84–86, 94, 123, 158
- Extrinsic Semiconductor
 - Definition of, 179, 187

- Face Centered Cubic Lattice, 109–110, 112, 141
 - First Brillouin Zone of, 126
 - Miller Indices, 120
 - Selection Rules, 139–140
- Family of Lattice Planes, 119, 120, 127, 135, 138
 - Spacing Between, 121
- Faraday's Law, 196
- FCC Lattice, *see* Face Centered Cubic Lattice

- Fermi
 - Energy, 29–31, 33, 37, 163, 166, 182
 - Level, *see* Fermi Energy, 33
 - Momentum, 29
 - Occupation Factor, 28, 31, 184, 185
 - Sea, 29, 30, 35, 166
 - Sphere, 29, 35
 - Statistics, 24, 25, 27–30, 35, 37, 184, 188

- Surface, 29–31, 92, 163, 166, 237
- Temperature, 29, 30, 32
- Velocity, 29, 30, 32, 35, 37
- Wavevector, 29, 33
- Fermi Liquid Theory, 37
- Fermi's Golden Rule, 132–133, 135
- Fermi-Dirac Statistics, *see* Fermi Statistics
- Ferrimagnetism, 214, 216, 217, 232
- Ferromagnetism, 199, 212, 216–225, 229–232, 240–241
 - Definition of, 196–197
 - Hard, 224
 - Itinerant, 235–239, 241
 - Nagaoka-Thouless, 241
 - Permanent, 224
- First Brillouin Zone, *see* Brillouin Zone, First
- Form Factor, 143
 - of Neutrons, 136, 137
 - of X-rays, 136–137
- Fractional Quantum Hall Effect, 3
- Free Electron Theory of Metals, *see* Sommerfeld Theory of Metals

- g-factor
 - Effective, 176
 - of Electron, 33
 - of Free spin, 204
- Gecko, 54
- General Relativity, 14
- Glass, 59
- Group Velocity, 71, 75, 176
- Gutzwiller Projection, 243

- Hall Effect, 35, 36, 179, 187
- Hall Resistivity, 21–23, 25
- Hall Sensor, 22
- Harmonic Oscillator, 8, 72
- Heat Capacity, *see* Specific Heat, *see* Specific Heat
 - of Diamond, 8, 10–11
 - of Gases, 7, 23
 - of Metals, 17, 24, 26, 30–32
 - of Solids, 7–17
 - Debye Model, *see* Debye Model of Solids
 - Einstein Model, *see* Einstein Model of Solids
 - Table of, 8
- Heisenberg Hamiltonian, *see* Heisenberg Model
- Heisenberg Model, 211, 212, 214–217
- Heisenberg Uncertainty, 240

- Higgs Boson, 2
- High Temperature Superconductors, 141
- Hole, 175, 176, 187
 - Effective Mass of, 176–177
 - Mobility of, 179
- Hope Diamond, 174
- Hopping, 50, 89
- Hubbard Interaction, 236, 241
- Hubbard Model, 235–243
- Hund's Rules, 197–205, 209, 211
- Hydrogen Bond, 42–44, 54–55
- Hydrogenic Impurity, 181
- Hysteresis, 222–224

- Impurities, 179–187
- Impurity Band, 182
- Impurity States, 180–183
- Indirect Band Gap, 173
- Indirect Gap, 164, 171–172
- Indirect Transition, 171–173
- Insulator, *see* Band Insulator or Mott Insulator
- Integral
 - Nasty, 14
- Intrinsic Semiconductor, 186
 - Definition of, 179, 187
- Ionic Bond, 42–46, 49
- Ionization Energy, 44–46
 - Table of, 45
- iPhone, 2, 189
- Ising Model, 211, 215–217, 219, 227–231
- Itinerant Ferromagnetism, *see* Ferromagnetism, Itinerant

- Karma, i, iv
- Kinetic Theory, 19, 23, 25
- Klechkovsky's Rule, *see* Madelung's Rule

- Landau Fermi Liquid Theory, 37
- Laser, 189
- Lattice, 78–79, 85, 99–112
 - Definition of, 99–101
- Lattice Constant, 64, 85, 108, 110, 121, 122, 143
 - Definition of, 78
- Lattice Plane, 118
 - Family of, *see* Family of Lattice Planes
- Laue Condition, 132–135, 140, 149
- Laue Equation, *see* Laue Condition
- Laue Method, 140

- Law of Dulong-Petit, *see* Dulong-Petit Law
 Law of Mass Action, *see* Mass Action, Law of, 188
 LCAO, *see* Linear Combination of Atomic Orbitals
 Lenz's Law, 196
 Linear Combination of Atomic Orbitals, 49
 Liquid, 58, 147
 Liquid-Crystal, 59
 Lorentz Correction, 143, 144
 Lorentz Force, 20, 32, 33
 Lorentz-Polarization Correction, 143
 Lorenz Number, 24
- Madelung Energy, 45
 Madelung Rule, 198
 Magnetic Devices, 245
 Magnetic Levitation, 196
 Magnetic Susceptibility, 196, 204, 205, 207, 209, 231, 239
 Magnetism, 32–34, 37, 170, 174, 195–210
 Animal, 195
 Magnetization, 33, 196, 222, 227, 237
 Mass Action, Law of, 186–188
 Mean Field Theory, 227–233, 236–239
 Metal, 92, 96, 163, 165, 174
 Metal-Insulator Transition, 95
 Metallic Bond, 42–44, 54, 91
 Miller Indices, 119–123, 127, 138
 for FCC and BCC Lattices, 120
 Minimal Coupling, 203
 Mobility, 21, 179, 187
 Modified Plane Wave, 160
 Molar Heat Capacity, 7, *see* Heat Capacity
 Molecular Crystal, 58
 Molecular Field Theory, *see* Mean Field Theory
 Molecular Orbital Theory, *see* Tight Binding Model
 Mott Antiferromagnetism, 239–241
 Mott Insulator, 170, 174, 208, 211, 239–241
 Multiplicity, *see* Scattering Multiplicity
- n-Dopant, *see* Donor
 Néel state, *see* Antiferromagnetism
 Néel Wall, 219–222
 Nearly Free Electron Model, 153–160, 166–169
 Nematic, 59
 Neutrons, 131, 136, 141, 153, 213
 Comparison with X-rays, 137, 148
 Sources, 149
 Spin of, 137
 Newton's Equations, 68, 79, 177–179
 Noether's Theorem, 75
 Non-Newtonian Fluid, 60
 Normal Modes, 68, 71–72, 75
 Enumeration of, 71–72, 163
 Nuclear Scattering Length, 136, 137
- One Dimension
 Diatomic Chain, 77–86
 Monatomic Chain, 65–75, 90
 Tight Binding Model, *see* Tight Binding Model of One Dimensional Solid
 Optical Mode, 82, 86
 Optical Properties, 82, 171–174
 Effect of Impurities, 173
 of Impurities, 182
 of Insulators and Semiconductors, 171–172
 of Metals, 36, 172–173
 Orthorhombic Lattice, 106
- p-Dopant, *see* Acceptor
 p-n Junction, 191
 Paramagnetism, 209, 231
 Curie, *see* Paramagnetism of Free Spins
 Definition of, 196
 Langevin, *see* Paramagnetism of Free Spins of Free Electrons, *see* Paramagnetism, Pauli
 of Free Spins, 203–206, 208–209, 232
 of Metals, *see* Paramagnetism, Pauli
 Pauli, 32–34, 205, 207–209, 239
 Van Vleck, 205, 206
 Particle in a Box, 47, 190, 201
 Particle-Wave Duality, 131
 Pauli Exclusion Principle, 24, 27, 30, 45, 178, 200, 201, 236, 240, 242
 Pauli Paramagnetism, *see* Paramagnetism, Pauli, 37
 Peltier Effect, 24–26, 32
 Periodic Boundary Conditions, 12–13, 28
 Periodic Table, 42, 44–46, 179, 199
 Perturbation Theory, 154–155, 206, 239
 Degenerate, 155
 Phase Velocity, 71, 75
 Phonon, 72–75, 90, 96, 172
 Definition of, 72

- Spectrum
 - of Diamond, 126
- Pinning, 218–219, 223–225
- Plan View, 107, 109, 112
- Polarizability, 53
- Polymer, 59
- Positron, 175
- Powder Diffraction, 141–146, 149
- Primitive Basis Vectors, *see* Primitive Lattice Vectors
- Primitive Lattice Vectors, 99, 116
- Primitive Unit Cell, 104, 112
 - Definition of, 102
- Proteins, 147

- Quantum Gravity, 2
- Quantum Well, 190
- Quarks, ii, 2

- Raise
 - Steve Simon Deserves, i
- Rant, 2–3, 42
- Reciprocal Lattice, 68–70, 74, 75, 96, 115–123, 127, 132–135, 138, 153, 154
 - as Fourier Transform, 117–118
 - Definition of, 70, 115–116
- Reciprocal Space, 74–75
 - Definition of, 69
- Reduced Zone Scheme, 81, 84, 86, 94, 123, 153
- Reductionism, 2–3, 42
- Refrigeration, 196, 206
 - Thermoelectric, 25
- Renormalization Group, 2
- Repeated Zone Scheme, 159
- Resistivity
 - Hall, *see* Hall Resistivity
 - of Metals, 21
- Riemann Hypothesis, 14
- Riemann Zeta Function, 14, 17–18
- Rotating Crystal Method, 140
- Rydberg, 181, 188

- Scattering, *see* Wave Scattering
 - Amplitudes, 135–137
 - Form Factor, *see* Form Factor
 - in Amorphous Solids, 147
 - in Liquids, 147
 - Inelastic, 147
 - Intensity, 135, 137–138, 143, 145
 - Multiplicity, 142
 - Scattering Time, 19, 23, 35, 179
 - Schrodinger Equation, 3, 11, 41–42, 44, 48–51, 88–90, 96, 155, 161, 242, 243
 - Seebeck Effect, 25, 32
 - Selection Rules, *see* Systematic Absences
 - Table of, 141
 - Semiconductor, 165, 174, 179
 - Devices, 189–191
 - Heterostructure, 190
 - Laser, 189
 - Physics, 175–188
 - Statistical Physics of, 182–187
 - Simple Cubic Lattice, 106, 108, 109, 112, 119, 120, 122, 125, 127, 137–141
 - Spacing Between Lattice Planes, 121
 - Slater Determinant, 28, 242
 - Soccer, 58
 - Somerville College, 147
 - Sommerfeld Theory of Metals, 27–37, 41
 - Shortcomings of, 35
 - Sound, 11, 14, 64–65, 70–72, 75, 81–82
 - Spaghetti Diagram, 126
 - Spallation, 149
 - Specific Heat, 7, *see* Heat Capacity
 - of Diamond, 8, 10–11
 - of Gases, 7, 23
 - of Metals, 17, 24, 26, 30–32, 37
 - of One Dimensional Quantum Model, 72–73
 - of Solids, 7–17
 - Boltzmann Model, *see* Boltzmann Model of Solids
 - Debye Model, *see* Debye Model of Solids
 - Einstein Model, *see* Einstein Model of Solids
 - Table of, 8
 - Spin Stiffness, 220, 221
 - Spin-orbit, 42, 176, 200
 - Spontaneous Order, 197, 212
 - Squalid State, ii
 - Stern-Gerlach Experiment, 137
 - Stoner Criterion, 237–239
 - Stoner Ferromagnetism, *see* Ferromagnetism, Itinerant
 - Structure Factor, 118, 135, 137–140, 143, 145, 149
 - Superconductor, 205, 235
 - Susceptibility

- Electric, *see* Polarizability
- Magnetic, *see* Magnetic Susceptibility
- Synchrotron, 149
- Systematic Absences, 138–141, 149

- Tetragonal Lattice, 106
- Thermal Conductivity, 23–25
- Thermal Expansion, 52, 65
- Thermoelectric, 25
- Thermopower, 25, 32
- Tight Binding Model, 153, 163, 168–169, 235–236, 241, 242
 - of Covalent Bond, 47–52
 - of One Dimensional Solid, 87–96
- Time-of-Flight, 149
- Topological Quantum Field Theory, 2

- Unit Cell, 77–79, 85, 94, 101–112
 - Conventional, *see* Conventional Unit Cell
 - Definition of, 78, 101
 - Primitive, *see* Primitive Unit Cell
 - Wigner-Seitz, *see* Wigner-Seitz Unit Cell

- Valence, 22, 36, 92, 93, 96, 163, 174
- Valence Band, 163, 164, 173, 175, 179
- Van der Waals Bond, 42–44, 53–54, 58
- Van Vleck Paramagnetism, *see* Paramagnetism,
Van Vleck
- Variational Method, 48, 88
- Virtual Crystal Approximation, 189, 232

- Wave Scattering, 131–149
- Weiss Domain, *see* Domain
- Weiss Mean Field Theory, *see* Mean Field
Theory
- Wiedemann-Franz Law, 24, 25, 32, 37
- Wigner-Seitz Unit Cell, 103–104, 109, 110,
112, 124–125, 127
 - of BCC Lattice, 109
 - of FCC Lattice, 110
- Wikipedia, 1

- X-rays, 131, 136–137, 140–141, 147, 153
 - Comparison with Neutrons, 137, 148
 - Sources, 148

- Zeeman Coupling, 33, 203, 223
- Zeeman Term, 203
- Zeta Function, *see* Riemann Zeta Function
- Zone Boundary, *see* Brillouin Zone Boundary