UK Graduate Summer School in Condensed Matter Theory

Physics by the Lake

Cumberland Lodge, Windsor Great Park – July 30 - Aug 10 2018



Features core lecture courses in condensed matter physics:

- **Strongly correlated quantum systems (Dr Chris Hooley, St Andrews)**
- **Statistical mechanics (Dr Richard Blythe, Edinburgh)**
- **Electrons in solids (Dr Martin Lüders, Daresbury)**
- **Cold matter and quantum fluids (TBA)**

and additional courses on a range of applications such as:

- **Quantum information processing (Prof Andrew Fisher, UCL)**
- **Soft condensed matter (Dr Buddhapriya Chakrabarti, Sheffield)**
- **Physics of biological evolution (Dr Bartlomiej Waclaw, Edinburgh)**
- **Topological matter (Dr Sam Carr, Kent)**
- **Mesoscopic physics and quantum coherence (Dr Ed McCann, Lancaster)**

Plus tutorials, research seminars and the chance to meet like-minded PhD students in beautiful surroundings

The school is aimed at condensed matter theorists in their 1st or 2nd year of a PhD

It is also accessible to experimentalists with an interest in theoretical topics

All meals and accommodation are included

The full fee for the school is £1400 which should be covered by students' training budgets

Some bursaries may be available in individual cases

International students are welcome

Full details at physicsbythelake.org

Apply by 21st April 2018 at physicsbythelake.org

Organisers: Prof Graeme Ackland (School of Physics and Astronomy, University of Edinburgh); Dr Sam Carr (School of Physical Sciences, University of Kent)

IOP Institute of Physics Theory of Condensed Matter Group · CM-CDT · CDT-CMP · CCP Magnetism · CCP5 · UKCP

The daily routine

- Breakfast is served each day from 8am
- Lunch is served each day from 12:45pm, with the exception of Sunday when a packed lunch is provided
- Dinner is served each day at 6pm

Schedule of lectures, tutorials and seminars

<table>
<tr><th></th><th>9am</th><th>10am</th><th>11am</th><th>12pm</th><th>1pm</th><th>2pm</th><th>3pm</th><th>4pm</th><th>5pm</th><th>6pm</th><th>7pm</th><th>8pm</th></tr>
<tr><td>Mon 30th</td><td colspan="6"></td><td colspan="4">Arrival</td><td rowspan="11">Dinner</td><td>CL intro</td></tr>
<tr><td>Tue 31st</td><td colspan="2">SCQ</td><td rowspan="10">Break</td><td>STM</td><td rowspan="10">Lunch</td><td>CQF</td><td rowspan="5">Free Time</td><td colspan="3">T-SCQ</td><td>S-1</td></tr>
<tr><td>Wed 1st</td><td colspan="2">STM</td><td>ELS</td><td>SCQ</td><td colspan="3">T-STM</td><td>P-1</td></tr>
<tr><td>Thu 2nd</td><td colspan="2">CQF</td><td>STM</td><td>SCQ</td><td colspan="3">T-CQF</td><td>S-2</td></tr>
<tr><td>Fri 3rd</td><td colspan="2">ELS</td><td>CQF</td><td>SCQ</td><td colspan="3">T-ELS</td><td>P-2</td></tr>
<tr><td>Sat 4th</td><td colspan="2">STM</td><td>SCQ</td><td colspan="2">T-MIX</td><td colspan="3">Free Time</td><td></td></tr>
<tr><td>Sun 5th</td><td colspan="9">Free Day</td><td></td></tr>
<tr><td>Mon 6th</td><td colspan="2">QIP</td><td rowspan="5">Break</td><td>SCM</td><td rowspan="5">Lunch</td><td>DFT</td><td rowspan="5">Free Time</td><td colspan="3">T-QIP+DFT</td><td>S-3</td></tr>
<tr><td>Tue 7th</td><td colspan="2">SCM</td><td>QIP</td><td>DFT</td><td colspan="3">T-SCM+DFT</td><td>S-4</td></tr>
<tr><td>Wed 8th</td><td colspan="2">MES</td><td>SCM</td><td>TOP</td><td colspan="3">T-MES+TOP</td><td>S-5</td></tr>
<tr><td>Thu 9th</td><td colspan="2">BIE</td><td>TOP</td><td>TOP</td><td colspan="3">T-MIX</td><td>S-6</td></tr>
<tr><td>Fri 10th</td><td colspan="2">MES</td><td>BIE</td><td colspan="6">Departure</td></tr>
</table>

Key to lectures and tutorials

Lecture	Tutorial	Topic
CQF	T-CQF	Correlated Quantum Fluids
ELS	T-ELS	Electrons in Solids
SCQ	T-SCQ	Strongly Correlated Systems
STM	T-STM	Statistical Mechanics
QIP	T-QIP	Quantum Information Processing
SCM	T-SCM	Soft Condensed Matter
DFT	T-DFT	Density Functional Theory
MES	T-MES	Mesoscopic Physics
TOP	T-TOP	Topological Phases of Matter
BIE		Physics of Biological Evolution
	T-MIX	Wildcard tutorial

We consider the courses in the first week to be core material, and cover material that one would expect any aspiring theoretical condensed matter physicist to be aware of, even if it is not your specialized topic. The courses in the second week are more specialized, applying the core material from the first week to different subfields of condensed matter theoretical physics. This gives you a breadth of knowledge that you often don't appreciate working on a PhD thesis, as well as illustrating that the same concepts appear again and again in seemingly completely different areas.

While lectures guide you through a subject, you don't really understand anything until you've tried to do questions on it yourself. In the tutorials, the lecturers and tutors will help you with this, as well as discussing with you any topic of physics you wish. While most of the tutorial sessions have been given a topic as a guide, please feel free to work on whichever question sheets interest you the most. Ultimately, there are far more problems than you will possibly have time to answer in two weeks!

Finally, if there is a topic you are interested in that isn't covered, then ask around – if one of the staff knows about this topic and is willing, and enough students are interested, then we can schedule extra optional lectures on rainy days!

Poster Sessions

The evening sessions marked P-1 and P-2 on the schedule are poster sessions. As there are too many posters to see in one evening, you will be randomly divided into two groups. At the beginning of the poster session, you will be asked to make a 1-minute introduction to your poster in the lecture theatre, before we go to see posters and have some drinks. There will be a prize for the best poster.

Key to Seminars

In the evenings when there are no poster sessions, we have research seminars, indicated by S-n (n=1-6) on the schedule.

S-1	Graeme Ackland	Metallic Hydrogen: The Holy Grail of High Pressure Physics
S-2	Mike Payne	The story of CASTEP
S-3	Julie Staunton	Spin-orbit coupling effects on electrons, magnetic anisotropy and crystal field effects
S-4	Raul Santos	Symmetry fractionalization in quasi-one dimensional systems
S-5	Alexandre Zagoskin	Schrödinger's elephants and quantum slide rules
S-6	Mike Gunn	TBA

Core Lecture Courses

CQF: Cold Matter and Quantum Fluids

Course and lecture notes written by Derek Lee (Imperial)
Lectures presented by members of the cast

Quantum fluids are those many-particle systems in whose behaviour the effects of both the quantum mechanics and quantum statistics are important, which occurs at cold temperatures. The most important two examples are superfluids, such as liquid Helium, and superconductors. This lecture course will begin with the phenomenon of Bose condensation in an ideal Bose gas with interactions; explore why this is not a true superfluid, and go on to look at the role of interactions. It then proceeds to explore what is different when the particles are charged, and finally look at the BCS theory of superconductivity where one begins with fermions rather than bosons.

ELS: Electrons in Solids

Martin Lueders (Daresbury)

A quantitative understanding of bonding in condensed matter systems demands a solution of the many electron problem. This course will show how the many electron problem can be mapped onto single electron problems in an approximate way (Hartree and Hartree Fock approximations) and a formally exact way (density functional theory and the Kohn Sham equations). Further, some of the methodology used to solve the Kohn Sham equations in complex systems will be described. In the last part of the lectures, some examples will be discussed, which show how electronic structure theory was able to explain some selected phenomena.

Martin is a principal scientist at Daresbury Laboratories. His main research interests are in the field of material-specific electronic structure theory of correlated systems, including normal, superconducting and magnetic states. The aim is to develop and apply methods within the density functional theory (DFT) framework and beyond, which allow to study materials of current interest from first principles.

SCQ: Strongly Correlated Quantum Systems

Chris Hooley (St Andrews)

This course deals mainly with the influence of interactions on the electrons in materials. We begin with a review of second quantisation and the Fermi gas theory of metals, and then progress to Landau's Fermi liquid theory and the notion of quasiparticles. The effect of impurities on the Fermi liquid (including the Kondo effect) is discussed, and we then move on to consider how the Fermi liquid gives way to other phases as the interactions are increased, concentrating on the Stoner instability and the Mott insulator. We analyse the magnetism in the Mott insulating phase, developing the concept of spin waves. Finally, we make a survey of some experimental data on strongly correlated crystalline solids, giving basic interpretations in terms of the concepts developed in the course.

Chris is a senior lecturer at the University of St Andrews. He works on various topics in the theory of strong correlations, including non-Fermi-liquids, highly frustrated magnets, non-equilibrium atomic fluids, and vortex-mediated phase transitions.

STM: Statistical Mechanics

Richard Blythe (University of Edinburgh)

Statistical Mechanics aims to provide a macroscopic description of a physical system starting from knowledge of its microscopic properties. The methodology and techniques are widely used throughout condensed matter physics and are also today being applied to understand the dynamics of model ecologies, economies and societies. In these lectures, we will revisit the equilibrium properties of matter - such as phase transitions and universality - from the perspective of dynamics (as opposed to statics, as is typically done in undergraduate courses). Then we will examine successively further-from-equilibrium systems, ending with a discussion of fluctuations in driven systems, a subject currently generating considerable excitement in this field.

Richard is a Reader at the University of Edinburgh. Since his PhD days, he has been researching models and theories for nonequilibrium dynamical systems. Applications of these models include transport in biological systems, traffic flow, population dynamics and language change.

Applications Lecture Courses

BIE: The physics of biological evolution

Bartlomiej Waclaw (School of Physics, University of Edinburgh)

The aim of this course is to discuss applications of statistical and quantum physics to the theory of Darwinian evolution. Lecture 1 will introduce basic concepts (selection, mutation, genetic drift) and simple models (Moran process) and explain how the methods of statistical physics can be applied to solve these models and derive important and quite general results. In lecture 2 we will discuss a more advanced model - the quasispecies model - and show how it can be solved in some special cases by mapping to a quantum spin chain model. Lecture 3 will discuss the role of spatial structure in biological populations of constant size, and in populations expanding into a previously unoccupied territory. The course will also highlight physicists' contribution to biological evolution. Lectures will combine PPT slides and blackboard derivations.

Bartlomiej (Bartek) is a statistical/biological physicist from Edinburgh, and a holder of the Royal Society of Edinburgh/Scottish Government Personal Research Fellowship. He has published research articles on random matrix theory, complex networks, random walks, and spin glasses. More recently, he has been working on applications of statistical physics to biological evolution, in particular the evolution of antibiotic resistance, and cancer progression.

DFT: A Practical guide to Density Functional Theory

Martin Lueders (Daresbury)

Density functional theory is one of the key computational techniques available today in condensed matter physics. While the theoretical background was

introduced in the electrons in solids course in the first week, this short course will involve a practical session and will allow students to see for themselves how to run DFT code, and what you get out of it.

MES: Mesoscopic Physics
Edward McCann (University of Lancaster)

Mesoscopic physics is the name given to electronic behaviour in solid state nanostructures that are so small that their size is similar to relevant characteristic length scales. Examples of such length scales include the elastic mean free path (which governs the scale for ballistic transport), the phase coherence length (quantum interference effects), and the electronic wavelength (quantum confinement). The aim of this course is to describe key experimental transport phenomena including weak localisation, universal conductance fluctuations, Aharonov-Bohm oscillations, and conductance quantisation whilst giving an overview of theoretical methods such as the tight binding model, the Landauer-Büttiker formulism, scattering theory, and scaling theory.

Ed is a reader in physics at the university of Lancaster. His research lies in the area of theoretical condensed matter physics, and he is particularly interested in the electronic properties of nanostructures, graphene and other two-dimensional materials.

QIP: Quantum Information Processing
Andrew Fisher (London Centre for Nanotechnology, UCL)

Quantum Information Processing is one of the most exciting applications of modern quantum physics, and has become a flourishing interdisciplinary field in its own right. In this short course we will concentrate on some aspects of the subject most relevant to condensed matter systems. We will start by defining qubits and quantum gates, then introduce quantum operations as a model for the action of a quantum system in a noisy environment and the Kraus representation theorem which provides a composite way to represent them. Then we will move on to quantum error correction and its connection to classical codes, and briefly discuss the physics of two of the most important solid-sate qubits: impurity spins in semiconductors and superconducting circuits. Finally we will talk about two alternatives to the standard gate model of quantum computation that particularly lend themselves to solid-state systems: adiabatic quantum computation (and the related topic of quantum annealing), and the topological computation (and related topological codes).

Andrew is Professor of Physics in the London Centre for Nanotechnology and the UCL Department of Physics and Astronomy; formerly Junior Research Fellow at St John's College Oxford (1989-93), Postdoctoral Fellow at the IBM Zurich Research Laboratory (1991-92), and Lecturer in Physics at the University of Durham (1993-95). He is Director of the new EPSRC Centre for Doctoral Training in Delivering Quantum Technologies, starting in 2014.

SCM: Soft Condensed Matter
Buddhapriya Chakrabarti (Sheffield University)

This course deals with the physics of soft materials. As the name suggests these materials are soft to touch (e.g. jello, creams, pastes etc.) as opposed to hard ones (e.g. metals, alloys)which fall under the purview of "Solid State Physics". The important distinction between soft materials as opposed to their hard counterparts is that entropy and not internal energy dictates their equilibrium properties. Further these materials mostly comprise of organic molecules that interact weakly and as a result their properties are strongly influenced by thermal fluctuations, external fields, and boundary effects. This strong 'susceptibility' of soft matter leads to many fascinating properties. We will review a few generic features of soft materials, e.g. dominance of entropy, interplay between broken-symmetry and dynamic mode structure and topological defects that are common to such systems. The outline is as follows i) Introduction to soft condensed matter physics, (ii) Liquid Crystals and Polymers (iii) Fluid Membranes, (iv) Fluctuations and response of non-equilibrium soft systems.

Buddhapriya (Buddho) is a senior lecturer in the biological physics group at the University of Sheffield. His main research interests include soft condensed matter physics and biological physics.

TOP: Topological phases
Sam Carr (University of Kent)

The well-known Landau theory of phase transitions classifies phases of matter according to broken symmetries and local order parameters, such as solids that break translational symmetry, or magnets that break magnetic rotation symmetry. It has been long known that there are phases of matter that defy this classification — the quantum Hall state being the most obvious (but by no means only) example. With the discovery of topological insulators about 10 years ago, interest in this field has exploded, and we now know of many distinct phases of matter with no local order parameter, but instead characterised by a topological invariant. This short lecture course will focus mostly on non-interacting band theory, and introduce topological invariants, boundary states, and the bulk-boundary correspondence necessary to understand the modern topic of topological insulators. Other manifestations of topology in modern condensed matter physics will also be exposed, although not discussed in detail.

Sam works on the theory of strongly correlated systems, specialising in low-dimensional systems both in and out of equilibrium. He has worked in groups in the US, Italy and Germany, and since 2013 has been a lecturer at the University of Kent in Canterbury.

Seminars

Metallic Hydrogen: The Holy Grail of High Pressure Physics

Graeme Ackland (University of Edinburgh)

At sufficiently high pressure, the electrons in the hydrogen molecules are squeezed out of their covalent bonds to become metallic. The basic properties of metallic hydrogen were already well described by the early theories of quantum mechanics in the 1930. Since then various predictions suggest it might be a room temperature superconductor, and superfluid, and supersolid or a supersonic rocket propellant. None of these has been detected experimentally, yet. In this after-dinner talk, I will cover the theoretical and experimental physics underlying the search. I will also give some insights into how research physics actually works, a discussion which may continue less formally in the bar later.

Graeme is Professor of Computer Simulation in Physics at the University of Edinburgh. He is also one of the organisers of this years physics by the lake school.

The story of CASTEP

Mike Payne (Cambridge)

In this talk I shall talk about the rise of plane wave total energy pseudopotential calculations based on density functional and the role of CASTEP in establishing this methodology. I shall also describe the commercialisation of CASTEP and discuss some of the advantages and disadvantages of combining academic research with making money.

Mike is Professor of Computational Physics at the University of Cambridge, and a Fellow of the Royal Society. His research group focuses the development of new tools and technologies which are designed to be accessible to all researchers.

Spin-orbit coupling effects on electrons, magnetic anisotropy and crystal field effect

Julie Staunton (University of Warwick)

Magnetic materials are typically characterised by a set of the saturation magnetisation M, exchange A and anisotropy K constants combined into two important length quantities, an `exchange length' and a `domain wall thickness'. This talk will outline the relativistic generalisation of density functional theory and its electronic structure basis and show how it enables M, A and K to be calculated. The spin-orbit coupling influence on the electronic structure of materials will be described and its role in the origin of magnetic anisotropy, K, will be discussed in particular. The effects upon the exchange interactions within small clusters of magnetic atoms deposited on substrates or between magnetic impurities will also be described. In magnetic materials with 4f rare earth elements the strongly correlated f-electrons form a non-spherically symmetric charge and magnetisation density which interacts with the surrounding charge distribution. In the talk it will be discussed how this crystal field effect and

strong spin-orbit coupling lead to significant magnetic anisotropy in rare earth - transition metal magnets.

Julie is professor of physics at the University of Warwick. Her research involves electronic structure theory to describe phenomena in materials from 'first-principles' using high performance computing techniques. Projects in theoretical magnetism, metal and alloy physics and ab-initio studies of properties of strongly correlated electron materials.

Symmetry fractionalization in quasi-one dimensional systems.

Raul Santos (Cambridge)

Symmetry plays a fundamental role in the classification of different phases on matter. In one and quasi-one dimensional systems, a complete understanding of different gapped phases is possible based on the presence of symmetry. It has been realized that among these symmetry protected states, new forms of quantum matter appear, with possible applications to quantum computing. I will review some classical examples and unveil the existence of new states in one and quasi-one dimensions.

Raul is a postdoc in Cambridge, and a tutor at this years physics by the lake school.

Schrödinger's elephants and quantum slide rules

Alexandre Zagoskin (University of Loughborough)

Alex is head of the physics department at the Univeristy of Loughborough, and works in quantum engineering, quantum metamaterials, and quantum computing.

Title TBA

Mike Gunn (University of Birmingham)

Mike is professor of theoretical physics at the University of Birmingham. His research is in the field of many body quantum physics, where he tries to determine the patterns they form at low temperatures (where quantum mechanics dominates), which might be patterns in space (a simple example would be a crystal lattice) but usually have a pattern in a more exotic sense - in their velocities or correlations between pairs of the particles. Currently the rapid experimental developments in ultracold quantum gases are providing him (along with many other people!) several interesting theoretical problems.

Tutors

Raul Santos (Cambridge)

Raul is a postdoctoral assistant at the university of Cambridge, having previously worked at the Weizmann institute in Israel, and Stony Brook university in the US. His research is in strongly correlated systems, using topics such as field theory and Bethe ansatz. Recently, he has worked a lot on topological phases of matter in strongly correlated systems.

Miriam Marques (Edinburgh)

Miriam is a research fellow in the physics department at the university of Edinburgh. Her research areas include computational materials physics and materials under extreme conditions.

Physics by the Lake

Physics by the Lake is a national summer school for PhD students in condensed matter theory in the UK. This year will be the 20th occurrence of the school. Many of the lecturers were themselves students at the school in the past – ask them about it!

The name `Physics by the Lake' comes from the original location of the school, in Ambleside beside Lake Windermere, in which it ran for more than 10 years (aside from one year in Wales – but that is another story entirely!) The location where we held the school however was mothballed in 2010, and the school had to move. After a few itinerant years, the school settled in its current location in Cumberland Lodge, which is close to Virgina water -- it was quite a challenge to find a location that filled all the usual academic requirements as well as being near a lake!

Until this year, the school has been funded by EPSRC. However, EPSRC have now withdrawn direct support for the school as their training money is routed through universities and CDTs. We do have support from a number of sponsors however – CCPmag, CM-CDT, CCP5, CCP9, psi-k and UKCP – who we would like to thank.

BIE

The physics of biological evolution

Bartek Waclaw
bwaclaw@staffmail.ed.ac.uk

School of Physics and Astronomy, The University of Edinburgh

Lecture notes for the graduate school "Physics by the Lake 2018", Cumberland Lodge, Windsor Great Park

Introduction

Biological evolution, as the name may suggest, has been traditionally the domain of biologists. An archetypal physicist who carefully avoided biology all his adult life but still remembers (from school?) something about "Darwinian evolution" might ask: what could be less interesting to a physicist than examining the shape and size of beaks of Darwin's finches (see Fig. 1)? Well, probably not very much[1]. However, it it just enough to apply our famous ability of "idealising" things (leaving out unimportant details) and the same physicist will jump and begin filling pages and pages with equations aimed at explaining biological evolution. How is that possible? The reason is that biological evolution, when stripped off the biological jargon is what physicists are very familiar with: it is the process of finding the ground state (the best adapted organism) in a complex space of all possible states (organisms) with many local energy minima (partial adaptations). This lecture series is about various aspects of this problem.

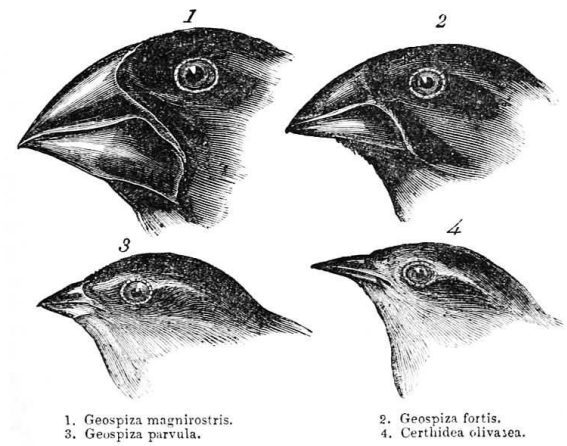


Figure 1: Darwin's finches by Gould (source: Wikipedia). Notice different shapes of beaks.

Although the physics of biological evolution is not even remotely as established as other branches of physics, and it is actually difficult to say whether such a sub-discipline exists on its own or is just a part of biophysics or statistical physics, it is worth to make here a short historical note on the contribution of physicists to biological evolution. The most known of them is probably Max Delbrück, a physicist-turned molecular biologist who (together with S. Luria, a microbiologist) demonstrated in 1943 that mutations occur at random with more-or-less constant rate and not, for example, as a response to selection[2]. Another example is Erwin Schrödinger with his "What is life" (1944), a book that has inspired many physicists and molecular biologists (including F. Crick and J. Watson who discovered the structure of DNA). Then, the late 70s saw M.

[1] Except perhaps for these researchers from Harvard: Campàs et al., PNAS **107**, 3356–60 (2012), see also `http://www.sciencedaily.com/releases/2010/02/100222182153.htm`.

[2] An opposite of that could be a controversial theory called "adaptive mutagenesis" which proposes that beneficial mutations occur in response to organism's altered needs, e.g., a change in the environment.

Eigen and P. Schuster (though not strictly speaking physicists but theoretical chemists[3]) and their quasispecies model aimed at explaining the evolution of self-replicating molecules. More recently, physicists[4] have contributed to a better understanding of the evolution of genetic regulatory/metabolic networks (U. Alon), fitness landscapes (J. Krug, V. Mustonen, M. Lässig), tumour evolution (A. Sottoriva), gene surfing, genetic drift in expanding populations (D. Nelson, O. Hallatschek), antibiotic resistance evolution (R. Austin, T. Hwa), and proposed various statistical models of evolving populations (E. Frey, T. Antal, R. McKane), to name only a few. The above list of problems and researchers is neither meant to be complete nor fully representative for the field which is still difficult to define.

So how can a contemporary theoretical physicist contribute to biology, and in particular to biological evolution? Traditionally, theoretical physicists are thought to have brought in a number of methods from statistical/quantum physics, but it is debatable how much of that was already known to mathematical/evolutionary biologists, perhaps under different names. Anyway, this belongs to the past and the only thing a physicist can do by carelessly re-discovering what these people already know is to annoy them (see e.g. [9]). Therefore I think that, apart from our most obvious strength - a general attitude that things can be explained and understood by modelling and determining what details are important and which are not - it pays off to do things that biologists (experimental and theoretical) do not usually do. For example, we can combine biological processes (growth) with physical processes (diffusion, flows, motility, biomechanics of cells). However, I have a few words of caution for anyone who, hearing for the first time about biological evolution during this lecture, embarks on a quest to formulate a biological "theory of everything":

1) Idealisation: one must be careful and not idealise too much (as in the joke about spherical cows in vacuum). Indeed, this is one of the sins that biologists accuse physicists of doing all the time: physicists look for universal laws where there are none and waste time on issues that have no biological meaning. After having listened to many biology-motivated conference talks given by physicists, it is hard for me not to agree with this statement. This can, and has done, some harm and, as a result, good physical biology that comes out of the physics community often goes unnoticed or is under appreciated by biologists.

2) Predictive models: a model that can explain all current experimental data is not worth much if it does not make any new predictions. This lack of predictability (and hence falsifiability) unfortunately plagues many models; usually it is not too difficult to construct an over-parametrized model that reproduces all the data but making new predictions is an entirely different story. However, there are also examples when modelling actually predicted new biology which has been confirmed since then. Some of such models (which, perhaps not surprisingly, have very few parameters) will be discussed in this lecture series.

3) Cross-fertilization between theory and experiment. This has always been the case in more traditional branches of physics, so why should it not apply to the physics of biological evolution? One problem that the reader may envisage is that biological evolution operates over the time-scale of millions of years, thus it cannot be replicated in the lab. However, this is true only when speaking about i) macroevolution, for example an ape evolving into a human, and ii) multicellular organisms which replicate slowly, say every 20 years. There are many examples of biological microevolutionary events which involve a few mutations only, happening over the time-scale of days, or even hours!

[3] Theoretical chemistry can in fact be regarded as physics in disguise (apologies if this offends any chemists!).
[4] Here: people employed in physics departments, educated as physicists, or using the methods of physics.

We are now ready to see examples of applications of the methods of physics to biological evolution. The first lecture will focus on the basics of evolution and translating these ideas into the language of physics. The role of physicists in what will be covered there is minimal although the methods used are the same as in statistical physics. Lectures 2 and 3 will provide more concrete examples of physicists' contribution to the theory of biological evolution.

The lecture notes cover a bit more than will be discussed during the actual lectures. This is to give those of you with a particular interest in biological evolution an opportunity to taste a wider selection of topics/methods, while simultaneously not overloading the rest. I will sometimes refer to the notes during the lectures to point to a derivation that I won't present on the blackboard, or an interesting problem I won't have time to discuss in full.

After each lecture there are two exercise problems: an easy one and a medium/difficult one. I do not expect that you solve all the problems, but I would like each of you to try to solve at least one of the total six.

1 Lecture. The basics of biological evolution

Some biological vocabulary

basepair (bp) A pair of nucleotides in double-stranded DNA; the smallest "bit" of genetic information. There are four possible pairs (A-T, T-A, G-C, and C-G).
point mutation The simplest heritable change in the DNA occurring when a basepair becomes replaced by another one either due to errors in DNA replication or chemical/physics agents (cancirogenes, X-rays etc.)
gene A fragment of DNA that codes a protein or an RNA chain that has a specific function in the organism. Genes have lengths ranging from 100bp to over 2Mbp.
allele One of a number of alternative forms of the same gene.
genotype The genetic make-up (DNA sequence) of an organism.
genotype space The space of all possible genotypes within a certain class of organisms.
genetic drift The change in allele frequency due to random variations in the number of organisms carrying that allele.
phenotype Observable traits of an organism, e.g. the size. The phenotype is determined not only by the genotype but also by environmental factors and the history of the organism.
fitness The measure of reproductive success of an organism. Here used interchangeably with growth rate.
fitness advantage A positive difference between the growth rate of a given organism and another one to which it is compared.
neutral No fitness advantage i.e. the growth rates are the same.
fitness landscape The genotype space, with genotypes connected by possible mutations, plus the growth rates (fitnesses) of all these genotypes.
selection The preferential survival of organisms with higher fitness.
fixation A change when only one allele or genotype remains in the population out of a few initially present.
speciation A process in which a population splits into genetically different groups whose members (organisms) cannot mate with other groups.
microevolution A change in alleles frequencies in a population whose members can mate with each other (as opposed to macroevolution which includes speciation). Occurs on relatively short time scales, down to a few h for microbes.

Evolution operates by two primary processes: natural selection, in which best adapted organisms outcompete less adapted ones, and genetic drift which is the source of variation in relative frequencies of different traits within the population. Although natural selection was supported by rather strong experimental evidence already 150 years ago, it was only in 1950's when the discovery of the role of DNA in heredity and, subsequently, the explanation of its structure provided a molecular basis for genetic variability. Today we know that this variability is caused by changes to the genetic material stored in the DNA, either by DNA exchange in sexual reproduction or microbial conjugation, or by mutations – random changes in the sequence of nucleotides which form the DNA.

In this lecture I will show how selection and genetic drift can be modelled mathematically. Strictly speaking, this has not been done by physicists but the methods used are that of dynamical systems and statistical mechanics. Throughout this and subsequent lectures we shall assume that we deal with a population of asexually reproducing organisms such as bacteria or viruses, which consume nutrients (glucose, aminoacids etc.) available in the environment and reproduce by making imperfect copies (due to mutations) of themselves. Such populations are easier to study mathematically than sexually reproducing organisms (such as *Homo sapiens*). We shall also assume that the organisms and nutrients are uniformly distributed in space and that the vessel in which growth takes place is being constantly stirred so that no gradients/inhomogeneities can occur. We shall call such models "well-mixed".

1.1 Genetic drift

The term "genetic drift" can be misleading to a physicist. Someone familiar with stochastic processes such as random walks or diffusion may think of "drift" as something (position of the population in some sort of phase space?) moving with a constant velocity. This is completely

opposite to the biological meaning of this term: it describes pure diffusion with no bias in the space of frequencies of different types of organisms (genotypes).

To understand genetic drift, we begin with a population of two variants (genotypes) A and B of a certain organism. A and B grow in the same environment and compete for common resources. We can think of B as a mutant version of A which differs only by a single point mutation[5] which has either just been produced from A, or evolved in a different spatial location and has been brought into contact with A (e.g., migrated or transported to the place where A lives).

We assume that the total size of the population is fixed and equal N. This simplifies calculations and is also not too far from some real-world experiments on microorganisms[39, 38]. We further assume that A and B reproduce with the same rate, and whenever one of them reproduces, a randomly chosen individual (A or B) is removed from the population to keep the total size constant. More precisely, at each time step a random individual is chosen for reproduction and a random individual is chosen for death. This model is called the Moran process (1958). The state of the system is fully described by just one variable, say the number i of organisms of type B. The role of the time variable is played in this model by the number n of update steps.

Figure 2 shows a computer simulation of this model, for a number of independent runs starting from the same initial condition $i = N/4$. A couple of interesting observations can be made:

- $i(n)$ strongly fluctuates before it eventually reaches either $i = 0$ or $i = N$. These two states are called "absorbing" because once either of them has been reached, the system stays there forever.
- More trajectories are absorbed at $i = 0$ than $i = N$ (true only for our specific initial condition).
- The time to absorption varies significantly (n_{final} between 200 and 900) among different trajectories.

Can we derive the above results mathematically? Fortunately, the Moran process is simple enough to be analytically solvable. We begin with listing the probabilities of all four events that can occur in a single time step:

- A is chosen for reproduction, and A is chosen for death. This happens with probability $((N-i)/N)^2$. After the event, i does not change.
- B is chosen for reproduction, and B is chosen for death. This happens with probability $(i/N)^2$. After the event, i does not change.
- A is chosen for reproduction, and B is chosen for death. This happens with probability $i(N-i)/N^2$. After the event, $i \to i-1$.
- B is chosen for reproduction, and A is chosen for death. This happens with probability $i(N-i)/N^2$. After the event, $i \to i+1$.

This can be rearranged for the probabilities of i staying the same or $i \to i \pm 1$. If we denote by M_{ji} the probability of going from state i to j, the only non-zero elements of this matrix are

$$M_{i+1,i} = M_{i-1,i} = i(N-i)/N^2, \quad (1)$$

$$M_{i,i} = ((N-i)/N)^2 + (i/N)^2 = 1 - M_{i-1,i} - M_{i+1,i}. \quad (2)$$

[5] Point mutations in which a single nucleotide has been replaced by another nucleotide are most common genetic alterations that occur in microevolution.

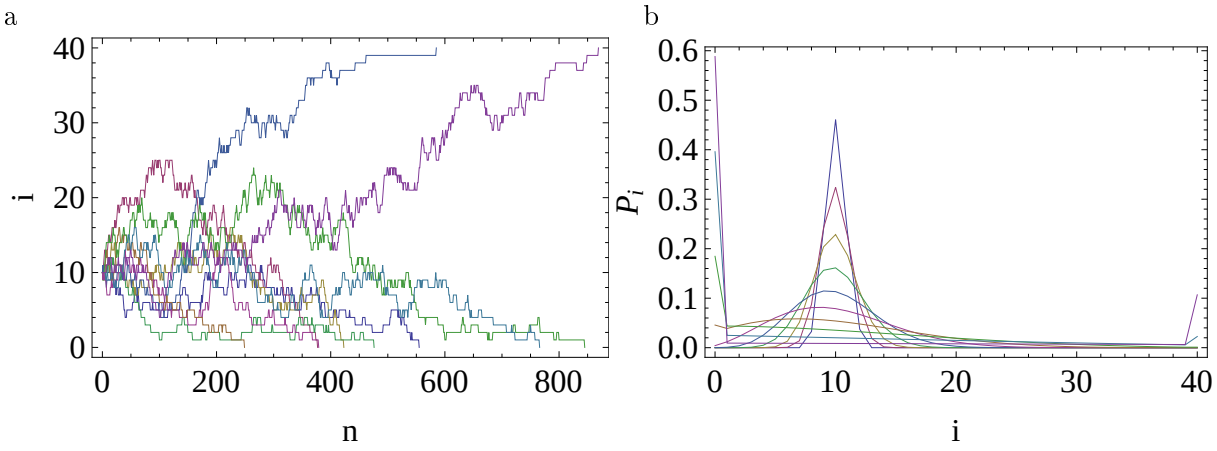


Figure 2: (a) 10 simulations of the Moran process for $N = 40$ and initial $i = 10$. (b) The probability $P_i(n)$ of finding i individuals of type B after n steps from Eq. (4) for $n = 1, 2, 4, \ldots, 1024$. The highest peak at $i = 10$ is for $n = 1$.

The matrix $\hat{M}$ is the Markov matrix of our stochastic Moran process, and defines a Markov chain:

$$P_i(n+1) = \sum_j (\hat{M})_{i,j} P_j(n), \tag{3}$$

where $P_i(n)$ is the probability of finding i organisms of type B after n steps. Iterating (3) we obtain that

$$P_i(n) = \sum_j (\hat{M}^n)_{i,j} P_j(0), \tag{4}$$

where $\hat{M}^n$ is the n-th power of the matrix $\hat{M}$. The states $i = \{0, N\}$ are absorbing because $M_{0,0} = M_{N,N} = 1$ and the system cannot escape from these states. This means that for sufficiently large number of steps the population will contain either only A or only B. Figure 2b shows $P_i(n)$ for a few different n and the initial condition $i = N/4$ as in Fig. 2a. According to our expectations, the probability becomes localized at the absorbing states $i = 0$ and $i = N$ for large times. Though the distribution P_i can be found analytically (cf. Ref. [15], p.106), we won't derive it here because it has limited practical use and the derivation is quite long.

A particularly interesting quantity often calculated for evolutionary models such as the Moran process is the probability F_i that B will reach fixation (the process becomes absorbed at $i = N$) when starting from i. Clearly, $F_N = 1$ and $F_0 = 0$. For any other i we have the following equation (see Richard's course for examples of similar equations):

$$F_i = M_{i-1,i} F_{i-1} + M_{i+1,i} F_{i+1} + M_{i,i} F_i. \tag{5}$$

The equation has three terms corresponding to three paths the system can take when starting from i: $M_{i\pm1,i} F_{i\pm1}$ is the probability of going from i to $i \pm 1$ times the probability of being absorbed from this new state, and $M_{i,i} F_i$ is the probability of staying in the same state after a single time step, times the probability of being absorbed from there. If we use (2) and rewrite (5) as

$$M_{i-1,i}(F_i - F_{i-1}) = M_{i+1,i}(F_{i+1} - F_i) \tag{6}$$

and notice that $M_{i-1,i} = M_{i+1,i}$ from Eq. (1), we obtain that $F_i - F_{i-1} = \text{const}$. Together with $F_N = 1$ and $F_0 = 0$ this gives us

$$F_i = i/N. \tag{7}$$

This is a very important result, although it seems now rather obvious: in our model, individuals reproduce with identical rates, hence each of them has equal chance $(1/N)$ of reaching fixation. If the initial number of such individuals is i, the fixation probability must be i/N. This is often the case with simple mathematical models; once the calculation has been performed, we can see how to arrive at the solution with much less effort!

Why is Eq. (7) so important? It shows that neutral mutations, that is these which do not affect the growth rate, have a very small chance of fixation (the mutation being present in all organisms) in large populations. For a 1ml of liquid bacterial culture which can contain more than 10^9 bacteria, the probability that a newly created, specific mutation will fix is $F_1 < 10^{-9}$. On the other hand, when the size of the population becomes small, neutral mutations fix much easier. An important evolutionary event - a population bottleneck - occurs when the size is reduced either due to environmental of human-made events and neutral (or even mildly deleterious) mutations may then spread in the population because F_1 is large.

Another interesting quantity we can compute is the average time to absorption; in the language of biological evolution, this is the time it takes for either of the types A,B to reach fixation (all cells in the population are either of type A or B). We must however define what we understand by "time" in the Moran process, where only probabilities of various events, not their rates (probability per unit time) are specified and hence there is no parameter that could set the time scale. Let us therefore *define* one unit of time as N steps of the Moran process. This corresponds to a constant replication rate per organisms - each of N organisms reproduces on average once every time unit and its reproduction rate does not depend on N, which is a reasonable biological assumption for not-too-dense populations. The time to absorption T, measured in these new units, is then equal to the number of steps n to absorption divided by N: $T = n/N$.

We first calculate the number of steps n_i to absorption on either of the two boundaries ($i = 0$ or $i = N$) if the initial number of organisms of type B is i. Clearly, n_i must obey the equation

$$n_i = 1 + n_{i-1}M_{i-1,i} + n_{i+1}M_{i+1,i} + n_iM_{i,i} \tag{8}$$

with two boundary conditions $n_0 = n_N = 0$. Notice a similarity to Eq. (5) - the only difference is the "plus one" term on the r.h.s. of the equation. Following the method of Ref. [18], we first solve a "homogeneous" equation (without the extra unity) and we obtain that $n_i - n_{i-1} = \text{const}$. We then make the constant i-dependent (a similar trick to what one does when solving inhomogeneous differential equations) and upon inserting $n_i - n_{i-1} = c_i$ into Eq. (8) we get

$$0 = 1 + c_{i+1}M_{i+1,i} - c_iM_{i-1,i} = 1 + (c_{i+1} - c_i)\frac{i(N-i)}{N^2} \tag{9}$$

from which it stems that

$$c_i = C_1 - \sum_{j=1}^{i-1} \frac{N^2}{j(N-j)}, \tag{10}$$

and thus

$$n_i = C_2 + \sum_{j=0}^{i-1} c_j = C_2 + \sum_{j=0}^{i-1} \left[C_1 - \sum_{k=1}^{j} \frac{N^2}{k(N-k)} \right], \tag{11}$$

with two constants C_1, C_2 which can be fixed by the two boundary conditions:

$$n_i = \frac{i}{N} \sum_{j=0}^{N-1} \sum_{k=1}^{j} \frac{N^2}{k(N-k)} - \sum_{j=0}^{i-1} \sum_{k=1}^{j} \frac{N^2}{k(N-k)}, \tag{12}$$

leading after some algebra to

$$n_i = \sum_{k=1}^{N-1} \frac{iN}{k} - \sum_{k=1}^{i-1} \frac{(i-k)N^2}{k(N-k)}. \tag{13}$$

In particular, for $i = 1$ we have

$$n_1 = N \sum_{k=1}^{N-1} \frac{1}{k} \cong N \ln N + \ldots. \tag{14}$$

These formulas are exact and can be used for numerical calculations but they become impractical for large N. However, in the limit $N \to \infty$ we can replace summation by integration in (13) and we obtain

$$n_i \cong N^2 \left[-x \ln x - (1-x)\ln(1-x)\right] + O\left(N \ln N\right), \tag{15}$$

where $x = i/N$. The time to absorption is therefore[6]

$$T_i = n_i/N \sim N \left[-x \ln x - (1-x)\ln(1-x)\right]. \tag{16}$$

When $x = O(1)$ or, in other words, a finite fraction of the population is initially taken by B, the time to either extinction or fixation of B scales linearly with the population size N. If $i = 1$ that is initially there is only one B, $T_1 \sim \ln N$ increases only logarithmically. How is that possible? For just one B present at time zero, extinction ($i = 0$) is much more likely than fixation and cases when the system becomes quickly absorbed at $i = 0$ dominate the average time T_1.

We may now ask what is the absorption time conditioned on the fixation of B, i.e. the system ending at $i = N$? We use the same approach as before, but the equation for n_i,

$$n_i = 1 + n_{i-1}\beta_i + n_{i+1}\alpha_i + n_i(1 - \alpha_i - \beta_i), \tag{17}$$

contains now probabilities α_i, β_i for jumps $i \to i \pm 1$ conditioned on the system being absorbed only at $i = N$:

$$\alpha_i = \frac{M_{i+1,i}F_{i+1}}{M_{i+1,i}F_{i+1} + M_{i-1,i}F_{i-1} + (1 - M_{i+1,i} - M_{i-1,i})F_i}, \tag{18}$$

$$\beta_i = \frac{M_{i-1,i}F_{i-1}}{M_{i+1,i}F_{i+1} + M_{i-1,i}F_{i-1} + (1 - M_{i+1,i} - M_{i-1,i})F_i}, \tag{19}$$

where $F_i = i/N$. The probabilities in the above equations are calculated using the definition of conditional probability: the probability of event X conditioned on event Y is $P(X|Y) = P(X \text{ and } \mathrm{Y})/\mathrm{P(Y)}$. For example, X in Eq. (18) corresponds to moving from i to $i + 1$, and Y corresponds to moving anywhere and finally being absorbed at $i = N$. $P(X \text{ and } \mathrm{Y})$, the probability of moving $i \to i + 1$ and finally being absorbed at $i = N$, equals to $M_{i+1,i}F_{i+1}$. The probability $P(Y)$ is the sum of probabilities of all three possible processes ($i \to i$, $i \to i \pm 1$ and eventually being absorbed at $i = N$).

Following the same route as in Eqs.(8-11) we find that

$$n_i = C_2 + \sum_{j=1}^{i-1} \left(\prod_{k=1}^{j} \frac{\beta_k}{\alpha_k} \right) \left(C_1 - \sum_{l=1}^{j} \frac{1}{\beta_l} \prod_{k=1}^{l-1} \frac{\alpha_k}{\beta_k} \right). \tag{20}$$

[6] Compare this to Richard Blythe's lecture notes where a different method is used to find the time to absorption for the Moran process.

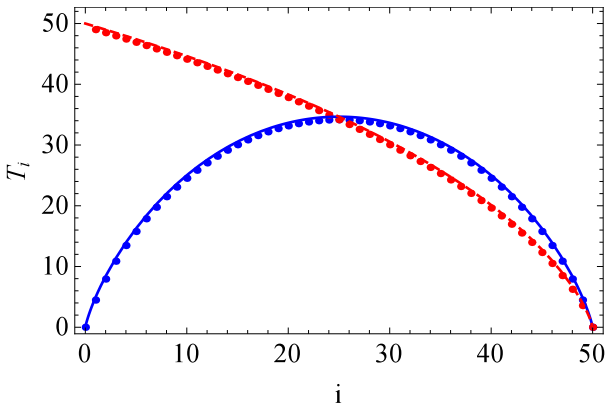


Figure 3: Average times to absorption (fixation) $T_i = n_i/N$ for the Moran process: absorption at either of the boundaries (blue, solid) and only at $i = N$ conditioned on the fixation of B (red, dashed). Points are from exact equations (13) and (23), lines are approximate formulas with $x = i/N$.

The constant C_2 is fixed by the boundary condition $n_N = 0$. To fix C_1, we note that n_i must be a monotonously decreasing, non-negative function of i. This leads to

$$n_i = \sum_{j=i}^{N-1} \sum_{k=1}^{j} \frac{1}{\beta_k} \prod_{l=k}^{j} \frac{\beta_l}{\alpha_l} = N^2 \sum_{j=i}^{N-1} \frac{1}{j(1+j)} \sum_{k=1}^{j} \frac{k}{N-k}. \tag{21}$$

n_1 can be expressed in a closed form by changing the order of summation:

$$n_1 = N^2 \sum_{k=1}^{N-1} \frac{k}{N-k} \sum_{j=k}^{N-1} \frac{1}{j(1+j)} = N^2 \sum_{k=1}^{N-1} \frac{k}{N-k} \frac{N-k}{Nk} = N(N-1), \tag{22}$$

which gives $T_1 = N-1$ - now the time to fixation of a single B scales linearly with N. Performing the same procedure for n_i for $i > 1$ we obtain:

$$n_i = N(N-i) \left[1 + \frac{1}{i} \sum_{k=1}^{i-1} \frac{k}{N-k} \right], \tag{23}$$

which, in the limit of large N and $x = i/N$ gives $T_i \cong N(1-1/x)\ln(1-x) + O(1)$. Figure 3 shows plots of all these fixation times.

To conclude our discussion of genetic drift, let me stress that some sort of randomness in how the organisms reproduce/die is absolutely crucial for this phenomenon. Although the second mechanism of biological evolution - selection (see next section) - can be understood using deterministic models (no randomness) described by ordinary differential equations for the average number of organisms of types A,B, genetic drift can not. To see this, let us consider the average value $\langle i(n) \rangle = \sum_i i P_i(n)$ of the number of B at time step n. Multiplying Eq. (3) by i on both sides and summing over i we obtain that

$$\langle i(n+1) \rangle = \sum_{i=0}^{N} i \left[2\frac{i(N-i)}{N^2} + \left(1 - \frac{2i(N-i)}{N^2} \right) \right] P_i(n) = \langle i(n) \rangle , \tag{24}$$

so the average i does not change in time at all! Had we not solved the stochastic case, we could have never guessed that there was something interesting (fixation/extinction) happening in the Moran process.

1.2 Selection

Natural selection is what happens when different organisms (different genotypes) have different reproduction rates. In contrast to genetic drift, randomness is not necessary to understand selection. We shall therefore begin with a simple deterministic model where stochasticity (and hence genetic drift) is absent. Let a, b be the average numbers of organisms of type A,B in a population of fixed size N. If A reproduces with rate r_A, and B with rate r_B, the time evolution of a, b is governed by

$$\frac{\mathrm{d}a}{\mathrm{d}t} = r_A a - D(t)a, \tag{25}$$

$$\frac{\mathrm{d}b}{\mathrm{d}t} = r_B b - D(t)b, \tag{26}$$

where $D(t)$ is such that the total size $a + b$ remains constant and equals N. Since constant size implies $\mathrm{d}a/\mathrm{d}t = -\mathrm{d}b/\mathrm{d}t$, we can compute $D(t)$ from the above equations as

$$D(t) = \frac{r_A a(t) + r_B b(t)}{N}. \tag{27}$$

This allows us to rewrite Eq. (26) as

$$\frac{\mathrm{d}b}{\mathrm{d}t} = b\left(1 - \frac{b}{N}\right)(r_B - r_A). \tag{28}$$

This so-called logistic equation can be solved exactly for $b(t)$ for any time, but instead of doing so we shall focus on the asymptotic behaviour for $t \to \infty$. The equation has two fixed points: $b = 0$ (extinction of B) or $b = N$ (fixation of B), but only one of them is stable for given r_A, r_B if the initial condition is $b(0) > 0$. If $r_A < r_B$, then $\mathrm{d}b/\mathrm{d}t > 0$ and b approaches N; the stable solution is thus $b = N$. Conversely, if $r_A > r_B$, $\mathrm{d}b/\mathrm{d}t < 0$ and b tends to zero; the stable solution is $b = 0$. This is the essence of selection: if organisms compete for resources (hidden in our model in the requirement that $a + b = N$), the fastest-growing organism dominates the population for $t \to \infty$. In the case when $r_A = r_B$ we have no selection and, as we know from previous section, deterministic equations cannot help us there.

Let us now discuss selection in finite, discrete populations. We shall again use the Moran process, with the following modification: A has fitness 1 whereas B has fitness $r \neq 1$. Both types have thus different reproduction rates per organism: 1 for A and r for B. To account for this in the Moran process, in each time step

- B is chosen for reproduction with probability $ri/(ri + N - i)$,
- A is chosen for reproduction with probability $(N - i)/(ri + N - i)$,
- B is chosen for death with probability i/N,
- A is chosen for death with probability $(N - i)/N$,

which gives the following transition probabilities $M_{j,i}$ for $i \to j$:

$$M_{i+1,i} = \frac{ri}{ri + N - i}\frac{N - i}{N}, \tag{29}$$

$$M_{i-1,i} = \frac{N - i}{ri + N - i}\frac{i}{N}, \tag{30}$$

$$M_{i,i} = 1 - M_{i-1,i} - M_{i+1,i}. \tag{31}$$

We want to find F_i, the probability of fixation of B when starting from i organisms of type B. Solving Eq. (5) in a similar way as before but for general $M_{i\pm1,i}$, and with boundary conditions $F_0 = 0$, $F_N = 1$, we obtain

$$F_i = \frac{1 + \sum_{j=1}^{i-1} \prod_{k=1}^{j} \gamma_k}{1 + \sum_{j=1}^{N-1} \prod_{k=1}^{j} \gamma_k}, \tag{32}$$

where $\gamma_k = M_{k-1,k}/M_{k+1,k}$. From Eqs. (29-30) we have that $\gamma_k = 1/r$ and therefore

$$F_i = \frac{1 + \sum_{j=1}^{i-1} r^{-j}}{1 + \sum_{j=1}^{N-1} r^{-j}} = \frac{1 - r^{-i}}{1 - r^{-N}}. \tag{33}$$

In particular, the fixation probability of a single B is

$$F_1 = \frac{1 - 1/r}{1 - 1/r^N} \tag{34}$$

which, for large populations $N \to \infty$ and for $r > 1$, reduces to

$$F_1 \cong 1 - 1/r \cong s, \tag{35}$$

where we have introduced the selective (fitness) advantage $s = r - 1$ of B over A. For large sizes and small (but not too small, $s \gg 1/N$) s, the fixation probability is thus equal to the selective advantage. This is one of the classical results of population genetics. It carries over to other models where populations have no spatial structure, and to many cases (but not all) of spatially-structured populations[7].

Table 1 gives F_1 for different r and N using Eq. (34) and the large-N form (35). We see that the approximate form works well for $N \gg 1/s$. Moreover, F_1 tends to the neutral-case expression $1/N$ when $s \ll 1/N$. This is an important observation - the size of the population must be much larger than $1/s$ to see selection, otherwise genetic drift will dominate even if the reproduction rates of A and B are different. The table gives also fixation probabilities for selective disadvantages $s < 0$ ($r < 1$) of the species B. The probability of fixation goes to zero quickly with increasing N even for very small selective disadvantage. This shows that deleterious genotypes with $s < 0$ fix in the population only if its size is much smaller than $1/|s|$. Large populations are therefore resistant to "backward evolution". This is not necessary always a good thing - imagine that in order to climb a higher peak in the fitness landscape the population has to go through a "fitness valley" where its fitness decreases before it can increase again. Equation (34) makes such situations extremely unlikely. They can occur nevertheless and this is called Muller's ratchet [32] - a process in which deleterious mutations can accumulate in asexually reproducing populations. This is one of the reasons why sexual reproduction is generally thought to be better for organisms.

To conclude this lecture let us estimate the rate R at which neutral mutations accumulate (fix) in the population. If a mutation arises with rate μ per organism and there are N organisms which replicate independently of each other, the rate of evolution $R = N\mu F_1$. For neutral mutations $F_1 = 1/N$ which gives $R = \mu$. The size dependence cancels out. This forms the basis of so-called "molecular clock" and in principle it allows biologists to tell how long ago two species must have separated from each other[8] by counting the number of neutral mutations that are different for

[7] We shall see examples of that in Lecture 3.

[8] This process is called speciation and involves a transition to the state where the population splits into two groups of organisms which cannot mate.

r	s	$F_1(N=\infty)$	$F_1(N=10000)$	$F_1(N=1000)$	$F_1(N=100)$	$F_1(N=10)$
0.5	-0.5	–	5×10^{-3011}	9.3×10^{-302}	7.9×10^{-31}	0.00098
0.9	-0.1	–	2.9×10^{-459}	1.9×10^{-47}	2.9×10^{-6}	0.059
0.99	-0.01	–	2.3×10^{-46}	4.4×10^{-7}	0.0058	0.095
0.999	-0.001	–	4.5×10^{-8}	0.00058	0.0095	0.1
1	0	–	0.0001	0.001	0.01	0.1
1.001	0.001	0.001	0.001	0.0016	0.0105	0.100
1.01	0.01	0.0099	0.0099	0.0099	0.016	0.104
1.1	0.1	0.091	0.0909	0.0909	0.0909	0.15
2	1	0.5	0.5	0.5	0.5	0.5

Table 1: Fixation probabilities for different selective (dis-)advantage s of species B and for different population sizes N. $F_1(N=\infty)$ has been calculated from Eq. (35) and $F_1(N=10,100,\dots)$ from Eq. (34). For the neutral case $s=0$ we use $F_1=1/N$.

the two species. This remains true also for size-changing populations. However, we just saw that the number of effectively neutral mutations can change as a function of N: mutations with small selective advantage can behave as neutral mutations for small N. Also, remember that we have assumed constant mutation rate μ while calculating R. This is not true for long periods of time - the mutation rate itself can evolve. So it is not clear at all how regularly spaced are the ticks of the molecular clock.

1.3 Further reading

All results presented in this lecture are in the books by M.A. Nowak [36] and W.J. Ewens [15]. For alternative derivation of the mean absorption time using generating functions see [34]. The Ewens' book contains also a detailed discussion of other evolutionary models (Wright-Fisher, the coalescent process, and others) and, like most textbooks on population genetics, deals also with sexual populations which we have not discussed here for the lack of time.

We have also completely neglected interactions other than competition between organisms, such as predation or cooperation. A large part of Nowak's book, written by a researcher who has made significant contributions to evolutionary game theory, is an excellent introduction to this problem.

Most evolutionary models assume populations of fixed size; populations whose size changes over time are much more difficult to analyse. See Refs. [2, 23] for an example of a model with mutation and selection in a growing population.

Exercise 1 (easy)

Assume we have only one species A which reproduces with rate b (probability per unit time) and dies with rate d, that is

$$A \to 2A \quad \text{with rate b,} \tag{36}$$

$$A \to 0 \quad \text{with rate d.} \tag{37}$$

Let there be only 1 A at $t=0$ and $b>d$. Show that (1) the average number of A increases in time as $\sim \exp((b-d)t)$, (2) the probability that A goes extinct (i.e. there remains no A after some time) is d/b.

Exercise 2 (medium)

Take the Moran model where B has fitness $r = 1 + s$ and A has fitness 1, and the population size is N. Physical time is defined as before: $T = n/N$ where n is the number of steps. There is only one B at $T = 0$.

1) How much time does it take (to the leading order) for species B to eliminate A from the population (i.e. fixation of B) if $s = r - 1 \gg 1/N$ and N is large?

2) How much time does that take (again, to the leading order) if $s < 0$?

3) Assume the population initially consists of A only. When A reproduces, it can mutate into B with some small probability $\gamma \ll 1$. If B has a selective advantage $s > 0$ and $\gamma N \ln N \ll s$, what is the average time it takes for B to reach fixation?

2 Lecture. The quasispecies model

In the previous lecture we discussed problems that, although having a flavour of statistical physics, were first solved by non-physicists[9]. We shall now discuss a model in which the methods of physics (and physicists) actually gave some new insight. This, and the fact that the authors of the model are physical chemists, must have contributed to the odium the quasispecies model faced from the biological community. There are papers (see e.g. [41, 35]) written by less-prejudiced mathematical biologists who try to explain the model to their peers but many population geneticists still refuse to accept it as a valid model of biological evolution, despite some successes the model has achieved in applications to real-world problems.

So what is this model that they fear so much? Remember the simple deterministic two-species model of selection from Lecture 1? The quasispecies is an extension of this model to the case of multiple species, plus the process of mutation in which one species can be transformed into another one. From now on we shall call these species genotypes. The model was initially conceived to mimic the evolution of self-replicating molecules such as RNA molecules thought to be present in the primordial "soup" in the oceans billions of years ago, or contemporary viruses such as HIV, but since then it has been used also for asexually-reproducing microorganisms such as bacteria. The name of the model comes from the fact that, in certain circumstances as we shall see below, the population forms a "cloud" of closely-related genotypes in the genotype space – a "quasispecies" – in analogy to the concept of the biological species viewed as a collection of closely-related organisms.

2.1 Model definition and properties

In the quasispecies model, each genotype is represented by a sequence of length L – the "genome" – composed of symbols taken from some finite set. The symbols can be the four nucleotides (A,T,G,C) which are the building blocks of DNA, but for simplicity one usually considers binary sequences composed of only two symbols 0 and 1. This assumption simplifies calculations and does not qualitatively change the properties of the model discussed in this lecture. The number of all possible "binary" genotypes is therefore 2^L. Genotypes are labelled by integers $i = 0, \ldots, 2^L - 1$, and the binary representation of i (padded with zeroes to keep the length constant) gives the corresponding binary sequence.

Organisms replicate by copying their genomes. However, the copy procedure is not error-free: each symbol (0 or 1) in the binary sequence has a finite probability γ of being substituted by the opposite symbol. Assuming that errors are made independently at each position in the genome (binary sequence), the matrix Q_{ji} which describes the probability of mutation from genotype i to genotype j reads

$$Q_{ji} = \gamma^{d(i,j)}(1-\gamma)^{L-d(i,j)}, \tag{38}$$

where $d(i,j)$ is the Hamming distance between genotypes i, j, i.e., $d(i,j)$ is the number of positions at which the corresponding sequences differ from each other. For $j = i$, $Q_{ii} = (1-\gamma)^L$ gives the probability of replication without errors. By definition, Q_{ji} is a symmetric, square, doubly stochastic matrix: $\sum_i Q_{ji} = \sum_i Q_{ij} = 1$.

[9] In the Pantheon of modern evolutionary theory, P. Moran and R. Fisher were statisticians, M. Kimura a mathematical biologist, J. Haldane a mathematician, T. Dobzhansky, S. Wright and J. Crow were geneticists. This list may, however, be misleading because it is often difficult to assign a single "profession" to those luminaries of science.

If we denote by r_i the specific growth rate of an organism with genotype i, we can describe the model by the following reactions

$$\text{genotype } i \xrightarrow{r_i} \begin{cases} \text{genotype } i + \text{genotype } i & \text{(with probability } Q_{ii}\text{)}, \\ \text{genotype } i + \text{genotype } j & \text{(with probability } Q_{ji}\text{)}. \end{cases} \tag{39}$$

where r_i above the arrow means that the reaction occurs with rate r_i. Suppose now that the population is very large so that we can neglect stochastic fluctuations of the numbers of organisms. Then, the time evolution of the numbers $n_0, \ldots, n_{2^L-1}$ of organisms of genotypes $i = 0, \ldots, 2^L-1$ can be described by the set of 2^L differential equations:

$$\frac{\mathrm{d}}{\mathrm{d}t} n_i(t) = \sum_{j=0}^{2^L-1} Q_{ij} r_j n_j(t) - n_i(t) J(t), \tag{40}$$

where $J(t)$ is the rate at which organisms are removed from the system. The term $J(t)$ forces the system to evolve towards the steady state, otherwise growth would always be exponential. The exact form of $J(t)$ is not important because it factors out in the analysis below, but for fixing our attention let us assume that $J(t)$ is proportional to the total population size, $J(t) = (1/K)\sum_i n_i(t)$, where K is some constant. This causes the net growth rates to become negative if the population is too dense, as if the organisms competed for limited resources.

Since $J(t)$ depends on $n_i(t)$, the quasispecies equation (40) is non-linear. However, a simple change of variables:

$$x_i(t) \equiv n_i(t) \exp\left(\int_0^t J(t')dt' \right) \tag{41}$$

reduces Eq. (40) to a linear equation,

$$\frac{\mathrm{d}}{\mathrm{d}t} x_i(t) = \sum_{j=1}^{N} Q_{ij} r_j x_j(t). \tag{42}$$

The above equation can be rewritten as

$$\frac{\mathrm{d}}{\mathrm{d}t} \vec{x}(t) = \hat{W} \vec{x}(t), \tag{43}$$

where the matrix $W_{ij} = Q_{ij} r_j$. Let $\lambda_0 > \lambda_1 > \cdots > \lambda_{2^L-1}$ be the set of eigenvalues of $\hat{W}$, and $\{\vec{w}_i\}$ denote their corresponding eigenvectors:

$$\hat{W} \vec{w}_i = \lambda_i \vec{w}_i. \tag{44}$$

For large times, equation (43) has the following asymptotic solution:

$$\vec{x}(t) \propto e^{t\lambda_0} \vec{w}_0. \tag{45}$$

Therefore, the steady state of Eq. (40) is proportional to the eigenvector $\vec{w}_0$ to the largest eigenvalue λ_0 of the matrix $\hat{W}$. This gives the steady-state numbers

$$\vec{n}^* \equiv \vec{n}(t \to \infty) = \frac{K\lambda_0}{\sum_i w_{0,i}} \vec{w}_0. \tag{46}$$

In this lecture we shall focus on the properties of steady-state solution (46) and, in particular, how they change with γ. We shall not discuss the dynamics of the quasispecies model - although

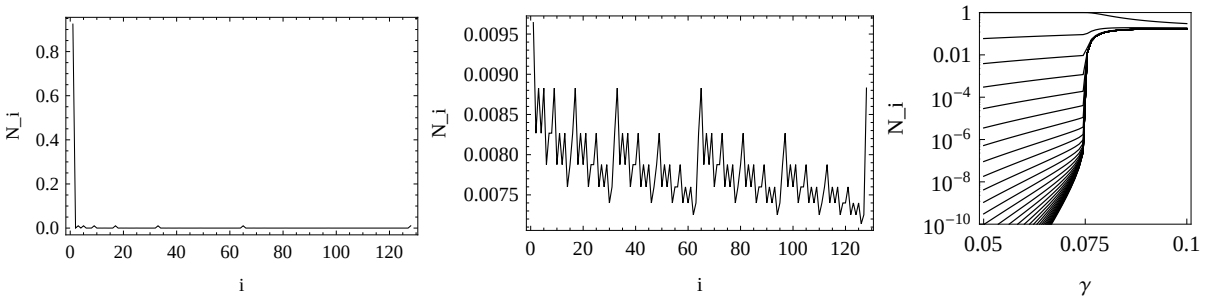


Figure 4: Fractions $n_i^*/n_{\rm tot}$ (where $n_{\rm tot} = \sum_j n_j^*$) of genotypes in the steady state in the single-peak model, for $L = 7$, $r_0 = 10, r_1 = \cdots = r_{127} = 1$, and for two mutation rates $\gamma = 0.01$ (left, well below the error threshold $\gamma_c(L = 7) \approx 0.28$) and $\gamma = 0.4$ (middle, above the threshold). Genotypes which correspond to the same Hamming distance from the master sequence $i = 0$ (the same "error class") have the same number of organisms. Right: plot of $n_i^*/n_{\rm tot}$ for different error classes $d(i, j) = 0, 1, \ldots$ (lines from top to bottom) as the function of γ, for $L = 30$. The transition, which is clearly visible at $\gamma_c(L = 30) \approx 0.074$, becomes even sharper for larger L.

very interesting, it is much more complicated to investigate than the steady state, and has been studied mostly numerically [26, 22].

The properties of the formal solution from Eq. (46) depend on the choice of the growth rates $\{r_i\}$. The space of all genotypes, connected by the graph of possible mutations (the matrix $\hat{Q}$), with fitness assigned to each genotype is usually referred to as the fitness landscape. This metaphor is based on viewing fitness peaks as mountains of different heights, with the population climbing generally uphill in the course of evolution and moving from lower to higher peaks, until the highest peak (global fitness maximum) is reached. Although frequently used in population biology, the concept of the fitness landscape has an important drawback: little information is available about real fitness landscapes because measuring the fitness experimentally for a large number of genotypes is very difficult. Based on available experimental evidence, many models have been proposed which catch certain aspects of the structure of real fitness landscapes. Here we shall consider just a few examples.

We shall begin with a single-peak landscape: the growth rate $r_0 > 1$ of the genotype $000\ldots00$ is taken to be maximal, and $r_1 = \cdots = r_{2^L-1} = 1$ are assumed to be smaller than r_0. This corresponds to a single best-fit genotype ("master sequence") and all mutants (other genotypes) being less fit. Figure 4 shows how $\{n_i\}$ change when the mutation probability γ is increased from a small to a large value. When γ is smaller than a certain critical value γ_c, the population is localized at the fittest genotype $i = 0$. Such a distribution of $\{n_i\}$ is called the quasispecies for reasons explained above. However, when $\gamma > \gamma_c$, the population becomes delocalized and the fraction of the population located at each genotype tends to $O(2^{-L})$ for $\gamma \to 1$. The critical γ_c is called the error threshold and the transition from localized to delocalized quasispecies is known as the error catastrophe because the increasing mutation probability (the probability of making errors during replication) causes the master sequence to vanish from the population.

The critical γ_c was first calculated by Eigen and Schuster in their seminal paper [13] using information theory, and then rederived by many others using different methods. A particularly simple derivation is presented in Ref. [36]. All these methods, however, rely on some hand-waving arguments that, though plausible, are not mathematically strict. Here we shall derive the expression for γ_c using an approach of Ref. [1]. The idea is to rewrite Eq. (44) as the Schrödinger equation for a quantum system whose ground state corresponds to the stationary solution - the

quasispecies. The ground state can then be found using methods of quantum mechanics, with all intermediate steps being either exact or involving well-controlled approximations. In addition, this approach works for several more complicated cases which cannot be easily tackled with other methods.

2.2 Quantum formalism for the quasispecies model

First, we notice that a binary sequence i of length L can be represented as a 1d spin chain with 1/2-spins $s_{i,1}, \ldots, s_{i,L}$ being either in the "+" state (which, by convention, we take to represent "0" in the binary sequence) or "−" (which represents "1"). Let us use Dirac notation for such a spin state:

$$|S_i\rangle = |s_i\rangle_1 \otimes \cdots \otimes |s_i\rangle_L = |s_{i,1} s_{i,2} \ldots s_{i,L}\rangle \,. \tag{47}$$

So, for example, for $L = 4$ the genotype $i = 0$ corresponds to $|S_0\rangle = |++++\rangle$, $i = 1$ to $|S_1\rangle = |-+++\rangle$, $i = 3$ to $|S_3\rangle = |--++\rangle$, and $i = 15$ to $|S_{15}\rangle = |----\rangle$.

The state of the system can now be represented by

$$|\psi\rangle = \sum_{i=0}^{2^L-1} x_i \, |S_i\rangle \,, \tag{48}$$

where $\{x_i\}$ are the same x_i's as in Eqs. (41,43). We can then rewrite Eq. (44) as the stationary Schrödinger equation

$$\hat{H} \, |\psi\rangle = E \, |\psi\rangle \,, \tag{49}$$

where $E = -\lambda$ and $\hat{H}$ is a quantum Hamiltonian operator that corresponds to the matrix $-\hat{W}$. How to determine this operator? We take the equation $\hat{W}\vec{x} = \lambda\vec{x}$, multiply it on both sides by $|S_i\rangle$, and sum over i:

$$\sum_i \sum_j Q_{ij} r_j x_j \, |S_i\rangle = \lambda \sum_i x_i \, |S_i\rangle \,, \tag{50}$$

which, after exchanging $i \leftrightarrow j$ on the l.h.s., becomes

$$\sum_i x_i \sum_j Q_{ji} r_i \, |S_j\rangle = \lambda \sum_i x_i \, |S_i\rangle \,. \tag{51}$$

The sums run over $0, \ldots, 2^L - 1$. To determine $\hat{H}$, we need to rewrite the l.h.s. as something acting on $\sum_i x_i \, |S_i\rangle \equiv |\psi\rangle$. Using the definition of Q_{ji} from Eq. (38) we have

$$\text{l.h.s.} = \sum_i x_i \sum_j \gamma^{d(i,j)} (1-\gamma)^{L-d(i,j)} r_i \, |S_j\rangle = (1-\gamma)^L \sum_i x_i \sum_j \left(\frac{\gamma}{1-\gamma}\right)^{d(i,j)} r_i \, |S_j\rangle \,. \tag{52}$$

Comparing this to the r.h.s. of Eq. (51) we see that we need two things: an operator that changes $|S_j\rangle$ into $|S_i\rangle$ and returns the factor $\left(\frac{\gamma}{1-\gamma}\right)^{d(i,j)}$, and an operator that returns r_i acting on $|S_i\rangle$. Perhaps not surprisingly (since we deal with spin-1/2 quantum chains) these operators can be constructed from Pauli spin operators:

$$\hat{\sigma}_k^z \, |\pm\rangle_k = \pm \, |\pm\rangle_k \qquad \text{and} \qquad \hat{\sigma}_k^x \, |\pm\rangle_k = |\mp\rangle_k \,, \tag{53}$$

where the subscript k means that the operator acts only in the Hilbert subspace of spin k. Although we are not going to use it, let us recall the matrix representation of these operators:

$$\hat{\sigma}_k^z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad \text{and} \qquad \hat{\sigma}_k^x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \tag{54}$$

so that when the states $|+\rangle = (1,0)^{\mathrm{T}}$ and $|-\rangle = (0,1)^{\mathrm{T}}$ are represented as vectors, these matrices fulfil equations (53). We can now define a new operator $\hat{R}$ whose eigenvalues are $\{r_i\}$:

$$\hat{R}(\hat{\sigma}_1^z, \ldots, \hat{\sigma}_L^z)\,|S_i\rangle = r_i\,|S_i\rangle\,. \tag{55}$$

It is clear that the operator does not have to depend on $\hat{\sigma}_k^x$; each $\hat{\sigma}_k^z$ returns the sign of the kth spin so if the operator depends on all $\hat{\sigma}_k^z$'s it can represent any (however complicated) fitness function.

The next step is to transform $|S_j\rangle$ into $|S_i\rangle$. We can do that using operators $\hat{\sigma}_i^x$ since they flip spins. We can write

$$|S_j\rangle = |s_j\rangle_1 \ldots |s_j\rangle_L = \prod_{k=1}^{L} (\hat{\sigma}_k^x)^{1-\delta_{s_{i,k},s_{j,k}}}\,|s_i\rangle_k = \left[\prod_{k=1}^{L} (\hat{\sigma}_k^x)^{1-\delta_{s_{i,k},s_{j,k}}}\right] |S_i\rangle\,. \tag{56}$$

We assume here that $(\hat{\sigma}_k^x)^0 = \nVdash_k$ is the identity operator, so the operator $(\hat{\sigma}_k^x)^{1-\delta_{s_{i,k},s_{j,k}}}$ flips the kth spin only if it is different in states i and j, thus the product over k transforms state j into state i. Therefore, the l.h.s. reads

$$\begin{aligned}
\text{l.h.s} &= (1-\gamma)^L \sum_{i=0}^{2^L-1} x_i \sum_{j=0}^{2^L-1} \left(\frac{\gamma}{1-\gamma}\right)^{d(i,j)} \left[\prod_{k=1}^{L} (\hat{\sigma}_k^x)^{1-\delta_{s_{i,k},s_{j,k}}}\right] R(\hat{\sigma}_1^z, \ldots, \hat{\sigma}_L^z)\,|S_i\rangle \\
&= (1-\gamma)^L \sum_{\{s_i\}} x_i \sum_{\{s_j\}} \left(\frac{\gamma}{1-\gamma}\right)^{L-\sum_{k=1}^{L}\delta_{s_{i,k},s_{j,k}}} \left[\prod_{k=1}^{L} (\hat{\sigma}_k^x)^{1-\delta_{s_{i,k},s_{j,k}}}\right] R(\hat{\sigma}_1^z, \ldots, \hat{\sigma}_L^z)\,|S_i\rangle \\
&= (1-\gamma)^L \sum_{\{s_i\}} x_i \left[\prod_{k=1}^{L} \sum_{\{s_j\}} \left(\frac{\gamma}{1-\gamma}\hat{\sigma}_k^x\right)^{1-\delta_{s_{i,k},s_{j,k}}}\right] R(\hat{\sigma}_1^z, \ldots, \hat{\sigma}_L^z)\,|S_i\rangle \\
&= (1-\gamma)^L \left[\prod_{k=1}^{L} \left(\nVdash_k + \frac{\gamma}{1-\gamma}\hat{\sigma}_k^x\right)\right] R(\hat{\sigma}_1^z, \ldots, \hat{\sigma}_L^z) \sum_i x_i\,|S_i\rangle\,.
\end{aligned} \tag{57}$$

In the above calculation we have replaced the sums over genotypes i, j with explicit sums over spins $\{s_i\}$ and $\{s_j\}$, each spin $s_{i,k}, s_{j,k}$ being either "+" or "−". By comparing the l.h.s. with the r.h.s of Eq. (51) we obtain that our Hamiltonian operator $\hat{H}$ reads

$$\hat{H} = -(1-\gamma)^L \left[\prod_{k=1}^{L} \left(\nVdash_k + \frac{\gamma}{1-\gamma}\hat{\sigma}_k^x\right)\right] R(\hat{\sigma}_1^z, \ldots, \hat{\sigma}_L^z). \tag{58}$$

Note that since $\hat{\sigma}_k^x$ and $\hat{\sigma}_k^z$ do not commute, the above Hamiltonian is in general non-Hermitian. However, this does not cause any problems for the ground state since we know from Perron-Frobenius theorem for our matrix $\hat{W} = -\hat{H}$ that it has a unique eigenvector to the largest eigenvalue λ_0, and that this eigenvector has non-negative entries, as the ground state of the Schrödinger equation should have.

Equation (58) is exact for any quasispecies model, but is difficult to analyse. To make progress, let us make two assumptions about fitness landscapes we are going to study. First, let the fitness (growth rate) depends only on the Hamming distance from genotype $|+++\cdots+\rangle$ ($i=0$). The operator $\hat{R}$ can then be written as

$$\hat{R}(\hat{\sigma}^z_1,\ldots,\hat{\sigma}^z_L) = \hat{R}\left(\sum_k \hat{\sigma}^z_k\right) = \hat{R}(\hat{\sigma}^z), \tag{59}$$

where we have used the notation $\hat{\sigma}^z = \sum_k \hat{\sigma}^z_k$ to denote the total z-spin operator. The form (59) includes the single-peak case from Fig. 4, and many others (two of them will be discussed below). Second, let $\gamma = q/L$ where q is independent of L. Since we saw that interesting things happened to the quasispecies model when γ was small, this seems to be a not-too-restrictive assumption. We want to know what happens in the large-L limit, that is for long binary sequences (large genomes). In this limit we have

$$\hat{H} = -\exp\left[-q + q\frac{\hat{\sigma}^x}{L}\right]\hat{R}(\hat{\sigma}^z) + O(1/L), \tag{60}$$

where $\hat{\sigma}^x = \sum_k \hat{\sigma}^x_k$ is the total x-spin operator. Now, since L is large and the Hamiltonian depends only on the x,z projections of the total spin operator, we expect the total spin to be large as well, and by invoking the correspondence principle[10] we can now replace quantum operators $\hat{\sigma}^x, \hat{\sigma}^z$ by their (classical) expectation values x, z and write[11]

$$H = -\exp\left[-q + q\frac{x}{L}\right]R(z) + O(1/L). \tag{61}$$

The ground state energy is given by the minimum of the above (now classical) Hamiltonian, for x,y,z constrained to the ball $x^2+y^2+z^2 \le L^2$ because at most L spins can align in the same direction. However, (61) is a monotonous function of x so any extremum of H must lie on the boundary of the ball. Since y does not appear in H, the extremum must be on the circle $x^2+z^2=L^2$. Replacing x by $\sqrt{L^2-z^2}$ in Eq. (61) we get

$$H(z) = -\exp\left[-q + q\sqrt{1-z^2/L^2}\right]R(z), \tag{62}$$

which needs to be minimized with respect to z to obtain the ground state energy.

2.3 Examples

Let us first discuss the single-peak case $r_0 = A > 1, r_1 = \cdots = r_{2^L-1} = 1$, for which $R(z) = 1 + (A-1)\Theta(z-L+\epsilon)$, with some small $\epsilon > 0$ and $\Theta(x)$ being the Heaviside (step) function: $\Theta(x) = 1$ for $x \ge 0$ and is zero for $x < 0$. The Hamiltonian reads

$$H(z) = -\exp\left[-q + q\sqrt{1-z^2/L^2}\right]\left[1 + (A-1)\Theta(z-L+\epsilon)\right]. \tag{63}$$

Figure 5, left, shows how this function looks like for $L=10, A=2$, and two different q. The function has two local minima,

$$H(z=0) = -1, \tag{64}$$

$$H(z=L-\epsilon) = -A\exp\left[-q + q\sqrt{1-(L-\epsilon)^2/L^2}\right] \cong -Ae^{-q}. \tag{65}$$

[10] We can do it because the Hamiltonian depends only on the total spin.

[11] For a more thorough discussion why this approach works see Ref. [1].

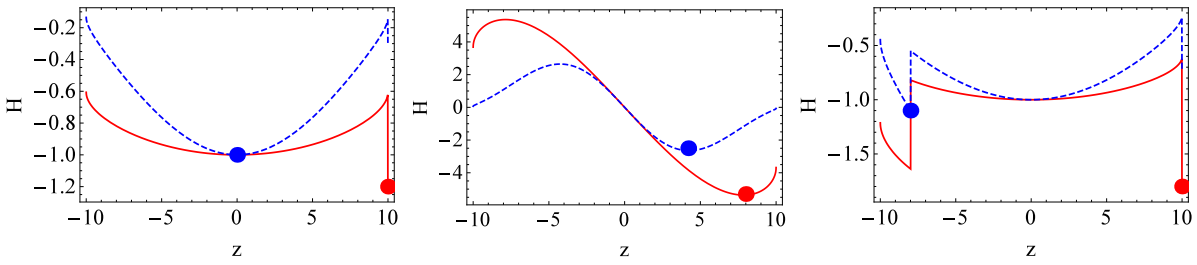


Figure 5: Left: $H(z)$ for the single-peak model and two different $q = 0.5$ (red, solid) and $q = 2$ (blue, dashed), for $L = 10, r_0 = A = 2$. Note the position of the global minimum (indicated by a small circle) shifts from $z = 9$ for $q = 0.5$ to $z = 0$ for $q = 2$. Middle: Mount Fuji fitness landscape, $L = 10$, for for $q = 1$ (red, solid) and $q = 5$ (blue, dashed). Right: two-peak landscape, $A_+ = 3, A_- = 2, k_+ = 1, k_- = 3, L = 10$, and for $q = 0.5$ (red, solid) and $q = 1.5$ (blue, dashed).

The first minimum at $z = 0$ is the global minimum (ground state) when $H(z = 0) < H(z = L-\epsilon)$ which leads to

$$q > \ln A + O(1/\sqrt{L}), \tag{66}$$

otherwise the second minimum is the global one. What do these two minima look like? Since z measures the average magnetization, $z = L - \epsilon$ means that almost all spins are up - the ground state is close to $|+ + \cdots +\rangle$ which is the fittest genotype $i = 0$. On the other hand, for $z = 0$ there must be on average equal number of spins up and down. This corresponds to a de-localized state, far away from the fittest one. The condition (66) gives thus the error threshold:

$$\gamma_c = \frac{\ln A}{L}, \tag{67}$$

which is the expression we sought for. Below γ_c, the population is localized around the fittest sequence, whereas for $\gamma > \gamma_c$ the population is delocalized. If we chose the fitnesses of unfit genotypes or $J(t)$ so that the population's average fitness (given by $\lambda_0 = -E_0 = -H(z = 0)$) was negative, the population could even become extinct in the limit $t \to \infty$.

One important thing that we have not discussed here because it involves the dynamics of the quasispecies model is that if γ is very small, evolution is very slow because it takes a lot of time for the population to move in the fitness landscape. It turns out that evolution is fastest just slightly below the error threshold and in fact it was found that fast-mutating viruses such as HIV operate close to γ_c [35].

Is the amazingly simple formula (67) the only outcome of our *tour de force* through the quantum formalism of the quasispecies model? Fortunately not. We can use Eq. (62) to derive some interesting results for other fitness landscapes. Let us first consider the case of additive fitness

$$R(z) = \alpha z, \tag{68}$$

which corresponds to $r_i = 2\alpha \sum_k s_{i,k} = 2\alpha(d(i,0) - L/2)$ (we assume here $s_{i,k} = \pm 1/2$). The fitness changes linearly with the distance from genotype $i = 0$. This kind of fitness landscape is called Mount Fuji landscape [4]. The minimum of

$$H(z) = -\alpha z \exp\left[-q + q\sqrt{1 - z^2/L^2}\right] \tag{69}$$

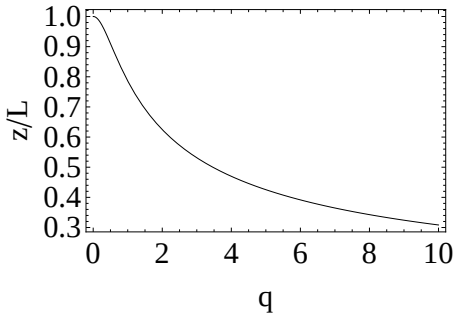


Figure 6: Plot of $z(q)/L$ for Mount Fuji landscape.

(cf. Fig. 5, middle) is at

$$z = L\frac{\sqrt{\sqrt{1+4q^2}-1}}{q\sqrt{2}}, \tag{70}$$

which is a monotonously decreasing function of q, see Fig. 6, and is always positive. There is no error threshold in this case; instead, the quasispecies distribution broadens and its centre shifts to genotypes with lower fitnesses as the mutation rate increases.

Lastly, we consider a two-peak landscape, with one sharp and high and one wide and low peak. Let the sharp peak of fitness A_+ and width k_+ be at $z = L$ and the lower peak with A_-, k_- at $z = -L$:

$$R(z) = 1 + (A_+ - 1)\Theta(z - L + k_+) + (A_- + 1)\Theta(-z - L + k_-), \tag{71}$$

with $A_+ > A_-$ and $k_+ < k_-$. The Hamiltonian has now three local minima, at $z = 0, z = L - k_+, z = -L + k_-$, see Fig. 5, right. Below the error threshold and when

$$\frac{A_+}{A_-} < \exp\left[q\left(\sqrt{2k_-/L} - \sqrt{2k_+/L}\right)\right], \tag{72}$$

the global minimum is at $z = -L + k_-$, that is the wider but lower peak is selected. This phenomenon is called the "survival of the flattest" and means that not only the height of the peak in the fitness landscape but also its width matter. Low, wide peaks may be selected over sharp and high peaks when the mutation probability is large enough. This can be explained as follows: when $q = \gamma L$ increases, the average fitness of a quasispecies located at a narrow peak goes down significantly because it is surrounded by genotypes with low fitness. On the other hand, a quasispecies residing at a lower but flat (wider) peak does not change its average fitness so much even if the quasispecies becomes slightly delocalized as the result of mutations.

2.4 Further reading

The molecular quasispecies theory has attracted considerable attention in the physics/theoretical chemistry community. Besides mapping it onto quantum spins discussed here [1], different versions of the quasispecies model has been studied by mapping it onto 2d Ising model [30, 10, 4, 5], directed polymers [27], or, more recently, Anderson localization [11, 14, 40]. Some exactly solvable special cases can be found in [4, 17, 42].

The quasispecies theory has been also generalized to the case when genes can be transferred "horizontally" between unrelated organisms [37] (this is a sort of sexual reproduction) and to spatially structured populations [40].

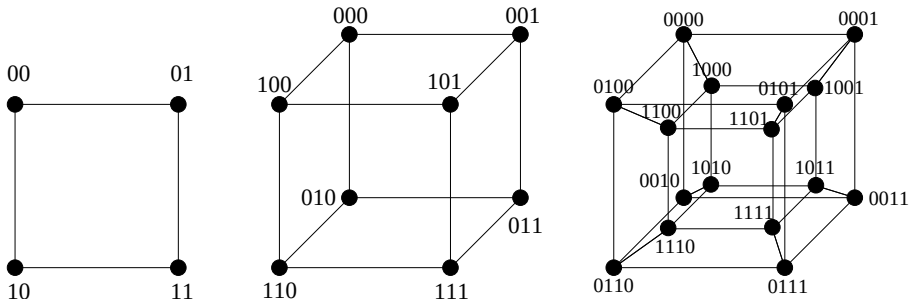


Figure 7: Examples of mutation graphs for the para-mu-se model, for $L = 2, 3, 4$.

Exercise 3 (easy)

Assume the same fitness for all genotypes. How does the ground state of such "neutral" quasispecies model ($R(z) = 1$) look like? What is the average magnetization z?

Exercise 4 (medium)

Consider the following alternative to the Eigen and Schuster quasispecies model:

$$\frac{\mathrm{d}}{\mathrm{d}t} n_i = n_i(r_i - J(t)) + \sum_{j=0}^{2^L-1} Q_{ij} n_j, \tag{73}$$

where the mutation matrix reads

$$Q_{ij} = \begin{cases} \gamma & \text{if } d(i,j) = 1 \\ -L\gamma & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{74}$$

and γ is now the mutation rate, not probability. In this model, mutations are allowed only along the edges of a hypercubic graph of dimension L, see Fig. 7. This model is called the para-mu-se model (parallel mutation and selection). Calculate the error threshold for this model, assuming as before that $r_0 = A > 1$ and $r_1 = \cdots = r_{2^L-1} = 1$.

3 Lecture. Spatially-structured populations.

In the last lecture we shall go beyond well-mixed (panmictic in the language of biologists) populations with no spatial structure and discuss how fixation of new genotypes is affected by spatial constraints. By this I mean that organisms which occupy distant spatial locations are isolated to a certain degree and, consequently, can create "patches" dominated by different genotypes in different locations. This is a typical biological scenario; mice that live in the UK are largely unaffected by what happens to mice in Germany, and even bacteria growing on one side of a Petri dish rarely bother what happens to their sisters on the other side. In fact, well-mixed populations are the exception rather than the rule, and rarely exist beyond the laboratory bench.

Many models of spatial populations have been proposed, each focusing on a different aspect of spatial growth. Here we shall discuss only two models: a spatial version of the Moran model, and a model of a population expanding into an empty 2d space. The latter is not only a nice illustration of how things change in spatially heterogeneous population. It is also an experimentally important case of growing bacterial colonies, and an example of physicists' contribution to biology which has been well received by biologists.

3.1 Spatial Moran model

In Lecture 1 we defined the Moran process by specifying the probability of choosing an individual for reproduction and for death. To model spatial populations, we must alter our definition slightly. However, we shall see that it will give exactly the same result in the well-mixed case where each organism can be replaced by any other organism.

In our new definition, we assume that individuals reside on nodes of a graph. A graph is a mathematical object composed of nodes and edges. The edges tell us which node is connected to which other node. If we label all nodes $i = 1, \dots N$, the graph is defined by an $N \times N$ matrix $\hat{A}$ called the adjacency matrix. The element A_{ij} equals one if there is an edge from node j to i, and is zero otherwise. Note that $\hat{A}$ does not have to be symmetric, that is in general edges do not have to be bi-directional; there may be an edge from j to i but not necessary from i to j. Such graphs which lack reciprocal edges are sometimes called "directed" as opposed to "undirected graphs" with symmetric $\hat{A}$. The total number of links coming out of node i is called the out-degree k_i of that node and reads $k_i = \sum_j A_{ji}$.

Examples of graphs are shown in Fig. 8. In particular, any regular lattice can be represented as a graph with a specific adjacency matrix, but graphs do not have to be regular, see Fig. 8d,e. Graphs that we shall consider in this section are[12]

(a) 1d periodic chain of length N: $A_{i,i+1} = A_{i+1,i} = 1$, the rest $A_{ij} = 0$, and we assume that $i = N + 1 \equiv 1$.

(b) 1d directed periodic chain of length N: $A_{i,i+1} = 1$, the rest $A_{ij} = 0$, and we assume that $i = N + 1 \equiv 1$. Note that $\hat{A}$ is asymmetric in this case.

(c) complete graph with self-loops: $A_{ij} = 1$ for all i, j.

(d) star graph: $A_{1,i} = A_{i,1} = 1$ for $i = 2, \dots, N$, the rest $A_{ij} = 0$.

[12] Mathematicians and physicists use different names even for most common graphs, so please bear that in mind when reading papers and books on evolutionary graph theory.

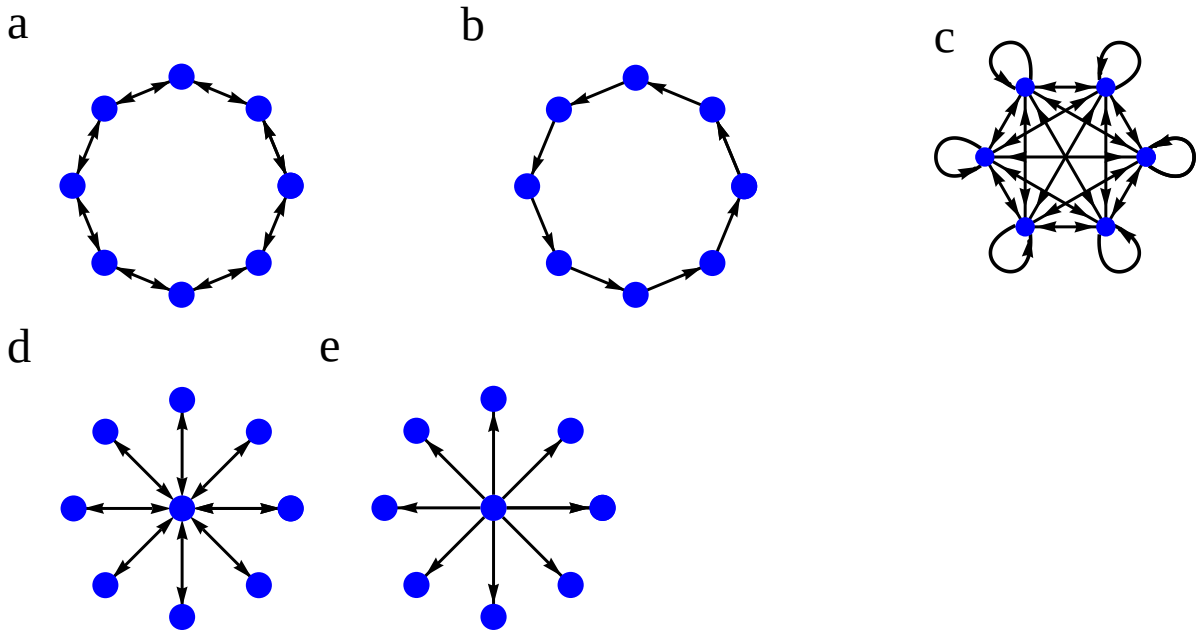


Figure 8: Examples of graphs: (a) 1d periodic chain, (b) directed periodic chain, (c) complete graph with self-loops, (d) star graph, (e) directed star.

(e) directed star graph: $A_{i,1} = 1$ for $i = 2, \ldots, N$, the rest $A_{ij} = 0$. This time the only edges come out from the central node and there are no reciprocal edges.

Let us come back to our new Moran model. Each node can be in one of two states A or B, depending on which organism lives on that node. As before, A has fitness 1, and B has fitness r. We denote the number of organisms of type B by m. At each time step we pick up a random node i with probability proportional to the fitness of an organism that resides there, so the probability is $(N-m)/(rm+N-m)$ for type A and $rm/(rm+N-m)$ for type B. This is exactly the same as before. Next, we choose a node j whose occupant will die and be replaced by a copy of the organism from node i. We assume that this happens with probability $w_{ji} = A_{ji}/k_i$, which means that it is equally likely to choose any neighbour of node i. The matrix $\hat{w}$ is stochastic: $\sum_j w_{ji} = 1$.

It is easy to show that the well-mixed Moran process from Lecture 1 is recovered for the complete graph with self-loops (Fig. 8c), for which $w_{ij} = 1/N$. Since each node is connected to any other node, the probability of selecting a node of type A for replacement is $(N-m)/N$, and a node of type B is m/N. From this we recover transition probabilities (29-31) (remember that i from those formulas corresponds now to m). The fixation probability F_1 of a single B is then given by Eq. (34):

$$F_1 = \frac{1-1/r}{1-1/r^N}. \tag{75}$$

What is the fixation probability for other graphs, for example 1d or 2d regular square lattices? It turns out that as long as the graph has the property that $\sum_i w_{ji} = \text{const} = 1$, fixation occurs with the same probability (75) as for the well-mixed Moran process. $\hat{w}$ must therefore be a doubly-stochastic matrix in which all rows and columns sum up to one. This statement is called the "isothermal theorem" by M.A. Nowak [36]. In our case where $w_{ji} = A_{ji}/k_i$, this is equivalent to saying that the graph is regular - all nodes have the same out-degree k. Since the 1d periodic chain, 1d periodic directed chain, 2d square lattice, and many other graphs are regular,

the fixation probability of the Moran process on these graphs is the same as in the well-mixed (complete graph) case.

It is straightforward to prove the isothermal theorem. Following [36], we define a vector $\vec{v} = \{v_1, \ldots, v_N\}$ describing the state of the system at time t. If node i is occupied by B then $v_i = 1$, otherwise $v_i = 0$. The total number of organisms of type B is then $m = \sum_i v_i$. At each time step, m increases by one with probability

$$M_{m+1,m} = \frac{r \sum_i \sum_j w_{ji} v_i (1 - v_j)}{rm + N - m}, \tag{76}$$

and decreases by one with probability

$$M_{m-1,m} = \frac{\sum_i \sum_j w_{ji} v_j (1 - v_i)}{rm + N - m}. \tag{77}$$

From Section 1.2 we know that F_1 assumes the form (75) if $\gamma_m = M_{m-1,m}/M_{m+1,m} = 1/r$. This is only possible if

$$\sum_i \sum_j w_{ji}(1 - v_i)v_j = \sum_i \sum_j w_{ji} v_i (1 - v_j), \tag{78}$$

which must hold for any vector $\vec{v}$ and, in particular, must hold for a vector in which $v_i = 0$ for all i except one $v_k = 1$ for some k. Inserting this vector into the above equation we obtain that, for any k,

$$\sum_i w_{ki} = \sum_j w_{jk} \tag{79}$$

and since $\sum_j w_{jk} = 1$ it must follow that $\sum_i w_{ki} = 1$ which completes our proof.

There are many graphs that do not fulfil assumptions of the isothermal theorem. We shall consider two examples here: a star graph, and a directed star graph. In both cases the central node has out-degree $N-1$ whereas peripheral nodes have out-degrees one (star) or zero (directed star), so the theorem does not apply. Let us first consider the directed star and assume that the first B is at a randomly chosen node. If the B happens to be at one of the peripheral nodes, it cannot reach fixation no matter how big r is because there is no reciprocal link through which they could invade other nodes (including the central node). The fixation probability must be then $F_1 = 1/N$ for any r because fixation can happen only if the first B is at the central node. The directed star completely eliminates selection and any new genotype behaves as a neutral mutant on that graph, irrespectively of its fitness advantage.

The (undirected) star graph, on the other had, amplifies selection. This means that for a given selective advantage $s = r - 1 > 0$ the fixation probability is larger than it would be for the well-mixed model with the same r,

$$F_1 \cong \frac{1 - 1/r^2}{1 - 1/r^{2N}}, \tag{80}$$

for sufficiently large (but finite) N. For derivation, see [8]. Figure 9 shows how this compares to the well-mixed F_1. It is possible to construct graphs that amplify selection even more, to the point that every genotype with arbitrarily small selective advantage $s > 0$ will reach fixation with probability 1, but we won't discuss them here. Examples of such graphs are in the book by M. A. Nowak [36] and a Nature paper by E. Lieberman [31].

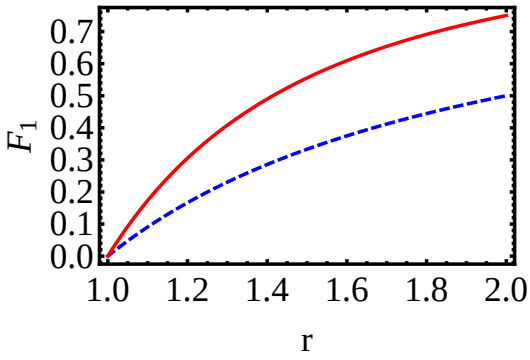


Figure 9: Fixation probability F_1 for the well-mixed Moran process (Eq. (75), blue, dashed) and the Moran process on the star graph (Eq. (80), red), in both cases $N \to \infty$.

3.2 Population expanding in 2d

We have seen that spatial structure can affect the fixation probability in comparison to the well-mixed case, but only if the population lives on a non-isothermal graph. This does not include regular lattices, so the Moran process in "normal" space (modelled by 2d or 3d lattice) does not differ from its well-mixed counterpart. But what if the population initially occupies only a fraction of space and expands into a previously empty territory as it grows in size (see Fig. 10)? This scenario occurs frequently in biology; examples range from bacterial colonies grown on Petri dishes, through microbial biofilms, to many invasive animal species including early human migrations.

Let us assume we have a population expanding vertically as in Fig. 10b with velocity v_A. The horizontal size of the population (the length of the moving front) is L. The population is composed of some tiny organisms but we do not model them explicitly here, we just assume that organisms with different genotypes form "sectors", i.e., spatial patches in which all organisms have the same genotype. Throughout this section we will be interested in the fixation probability of a new genotype B in the population of A. Let B has initially the horizontal extension w_0 when the front is at position y_0, which corresponds to time $t_0 = y_0/v_A$. The extension w_0 plays a similar role to the initial number of B-type organisms in the well-mixed model.

We first consider a sector formed by a neutral mutant B that expands with the same velocity $v_B = v_A$ as the "wild type" A. Figure 11, left, shows two possible fates - fixation and extinction - of that sector. We assume that boundary walls $x_-(y), x_+(y)$ of the sector can be described by Brownian motion (diffusion):

$$\frac{\mathrm{d}x_\pm}{\mathrm{d}y} = \sqrt{2D}\eta_\pm(y), \tag{81}$$

where D in the above Langevin equation is the diffusion constant of the boundary wall and $\eta_\pm(y)$ is uncorrelated random noise (see Richard's lectures) with zero mean and unit variance:

$$\langle \eta(y) \rangle = 0, \qquad \text{and} \qquad \langle \eta(y)\eta(y') \rangle = \delta(y - y'). \tag{82}$$

The idea behind using Eq. (81) to model the behaviour of sector domain walls is that if cells replicated randomly at the front, sectors harbouring different genotypes would wander randomly to the left or right and their behaviour would be described by random walk or, in the continuous limit, by Brownian motion.

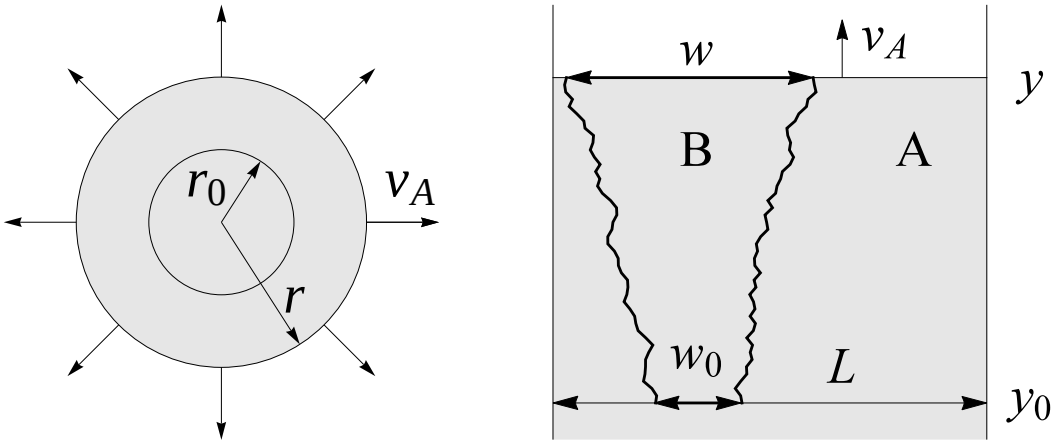


Figure 10: Populations in which organisms grow in a narrow layer at the front. The front moves with velocity v_A. Left: circular colony. Right: population expanding vertically in a tube of horizontal width L. A sector of type-B organisms expands from w_0 to w as the front moves vertically from y_0 to y.

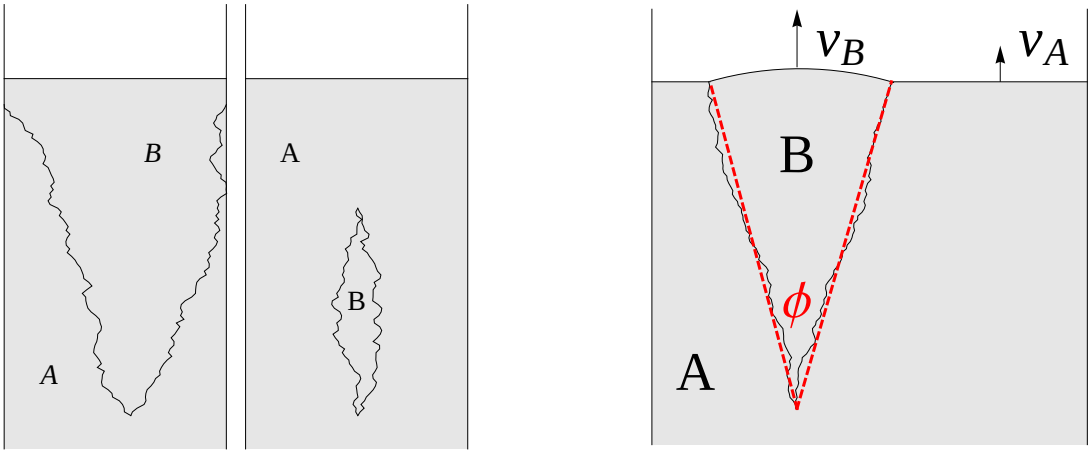


Figure 11: Left: Two possible fates of a B sector: fixation (the sector size $w = L$) and extinction ($w = 0$). Right: advantageous mutation spreading with velocity $v_B > v_A$ forms a triangular sector.

What is the horizontal size $w(y)$ of the mutant sector at position y? Obviously, $w(y) = |x_+(y) - x_-(y)|$ and hence $w(y)$ also performs Brownian motion,

$$\frac{\mathrm{d}w}{\mathrm{d}y} = \sqrt{4D}\eta(y), \tag{83}$$

albeit with twice as large diffusion constant (try to derive it!). Let us denote by $F(y|w_0, y_0)$ the probability that a sector of size w_0 at y_0 will reach fixation at y. It can be shown (cf. Richard's STM lecture) that $F(y|w_0, y_0)$ obeys the following backward equation,

$$\partial_{y_0} F(y|w_0, y_0) = -2D\partial_{w_0 w_0} F(y|w_0, y_0), \tag{84}$$

with two boundary conditions $F(y|0, y_0) = 0$ and $F(y|L, y_0) = 1$. Although it is possible to solve this equation for arbitrary y [21], we shall only look at the case of $y \to \infty$, i.e., very long times. In this case the l.h.s. of Eq. (84) is zero because the probability cannot depend on y_0 when $y - y_0 \to \infty$, and we obtain by integrating the equation twice that

$$F(\infty|w_0, y_0) = w_0/L. \tag{85}$$

This is the result we might have expected and it is a spatial counterpart of F_1 from Eq. (7). Neutral mutants thus fix with the same probability in the spatial, expanding population as in the well-mixed case. No surprise so far.

Now, let us calculate $F(\infty|w_0, y_0)$ for an advantageous mutant which forms a sector advancing with velocity $v_B > v_A$. Figure 11, right, shows how the sector looks like in this case. First, the boundaries are pulled apart because of the growth advantage of B over A. We can model the boundaries as trajectories of biased Brownian particles with the same diffusion coefficient D as before but with some small drift $m > 0$ (to be determined later) in two opposite directions. The sector size w then evolves according to the following Langevin equation:

$$\frac{\mathrm{d}w}{\mathrm{d}y} = \sqrt{4D}\eta(y) + 2m, \tag{86}$$

and the backward equation for $F(y|w_0, y_0)$ reads

$$\partial_{y_0} F(y|w_0, y_0) = -2D\partial_{w_0 w_0} F(y|w_0, y_0) - 2m\partial_{w_0} F(y|w_0, y_0), \tag{87}$$

with boundary conditions $F(y|0, y_0) = 0$ and $F(y|L, y_0) = 1$. Again, we solve it for $y \to \infty$ and we obtain

$$F(w_0) \equiv F(\infty|w_0, y_0) = \frac{1 - \exp(-mw_0/D)}{1 - \exp(-mL/D)}. \tag{88}$$

This equation is formally equivalent to Eq. (34) if we take $r = \exp(mw_0/D)$ and $N = L/w_0$. Again, nothing surprising (or so it seems!). But there is still one step to make: we must relate m to the expansion velocity v_B of the type B.

Let us look again at the sector drawn in Fig. 11, right. It has a triangular shape and is capped with an arc of a circle. How do we know that? In our model, the population expands in the direction normal to the front. If the population is composed of just one type of organism and if it starts from a flat front, the population will simply expand vertically. We can view this as each point on the surface creating a circle of radius $v_A dt$ in the infinitesimally small time dt, and the front at time $t + dt$ being the outer envelope of such circles [3]. This is similar to the Huygens principle for waves. Now, an infinitesimally small mutant sector in the absence of Brownian motion will create a circle of a slightly larger radius $v_B dt$. The mutant sector and the rest of

the front must be glued together at the boundaries, hence $v_A dt = \cos(\phi/2) v_B dt$ where ϕ is the opening angle of the sector. During the time dt the mutant sector will spread horizontally by $v_B \sin(\phi) dt$, hence our drift velocity m (measured in units of length/length) is

$$m = \frac{v_B}{v_A} \sin(\phi) = \frac{v_B}{v_A} \sqrt{1 - \frac{v_A^2}{v_B^2}} = \sqrt{\frac{v_B^2}{v_A^2} - 1}. \tag{89}$$

If we now take $v_B = v_A(1+s)$ where s is the selective advantage of genotype B, we obtain that

$$m = \sqrt{(1+s)^2 - 1} \cong \sqrt{2s} \tag{90}$$

for small s. Inserting this into Eq. (88) gives, in the limit of large L,

$$F(w_0) \cong \frac{m w_0}{D} \cong \frac{w_0 \sqrt{2s}}{D}. \tag{91}$$

The fixation probability is no longer proportional to s as it was in the well-mixed case (35) but to the square root of s. This means that it is easier for a mutant with a small fitness advantage to fix in the spatially expanding population that in the population of fixed size (well-mixed or not). Second, the probability is inversely proportional to the wall diffusion constant D which depends on the microscopic details of cell-cell biomechanical interactions. We shall come back to this later.

How sensitive is formula (91) to various assumptions we have made in this model? For example, what if the front remains flat in the presence of faster-growing mutants B who can only expand horizontally? Although this is not very likely to happen in reality (in bacterial colonies one always sees funnel-like sectors with curved caps as in Fig. 11), let us briefly discuss it. If some phenomenon keeps the front flat but at the same time does not change the growth rate of B, biomass conservation imposes that m must be proportional to s (new bacteria must go somewhere). In this case, $F(w_0) \propto (w_0/D)s$ scales linearly with s as in the well-mixed Moran process. In fact, a 1+1d Moran process in which a cell in a new line of cells inherits the genotype of one of its three neighbours in the previous layer, and the lines are stacked vertically to form a 2d plane, gives $F_1 \cong 2s$. This can be easily understood: the size of the sector can be mapped to the number of B in the well-mixed Moran process with twice the fitness advantage because the sector can grow in both directions (left and right).

The second, and more serious problem is how good is our approximation of sector boundaries by Brownian motion? Shall we not, instead of normal diffusion, consider superdiffusion for which

$$\left\langle (x(t) - x_0)^2 \right\rangle \propto t^{2\zeta} \tag{92}$$

with $\zeta > 1$? For example, a simple lattice model of spatially expanding populations called the Eden model [12] predicts $\zeta = 2/3$ [20], and this seems to agree with experimental results for bacterial colonies on Petri dishes [20]. Although the exponent $\zeta = 2/3$ does not differ very much from that of Brownian motion ($\zeta = 1/2$), it can create a perceivable difference in $F(w_0)$ for large times and small initial sizes of the sector. Moreover, non-Brownian behaviour of $x(t)$ is known to occur in situations in which the growing front is not flat (even in the neutral case) and so we cannot guarantee that Eq. (89) still holds because it was derived assuming locally flat fronts. However, experimental evidence suggests that although formula (91) may not be universal, spatial expansion generally changes the fixation probability as compared to Eq. (75).

3.3 Wall diffusion constant

We have not spoken too much about what determines D and the reason is that this is a very complex problem. Recent research suggests that in microscopic models of bacterial colonies, in which microorganisms are represented as particles interacting mechanically, consuming nutrients diffusing in the substrate on which they live, growing and dividing when reaching some critical size,

- D is mostly determined by "roughness" of the growing front [19]. If the front is almost flat (it is never completely flat due to intrinsic discreteness of cells), D is small. If the front is very rough (there are large fluctuations of the front height y as function of x), D is large.
- front roughness is larger for lower nutrient concentration, which promotes "bulging" of the front as a mean of gaining better access to food [16].
- the shape of cells affects D. Generally, the less regular cells are (less spherical), the larger D is.
- mechanical interactions between cells, and between the cells and the surface on which they live, affect D. For example, depending on whether cell-surface friction is isotropic or anisotropic, D can assume different values.

Experimental evidence [19] shows these findings obtained in computer simulations are at least qualitatively correct, but there is still much to be done before we can say we understand how fixation is affected by mechanical properties of cells in expanding populations.

3.4 Further reading

Martin Nowak's book [36] contains several chapters on evolutionary theory and spatial games, and is a good starting point for anyone interested in evolutionary dynamics. Research on expanding bacterial colonies is relatively new (last 8 years or so). Papers by O. Hallatschek and D. Nelson [20, 21] introduce basic models discussed here and make qualitative comparisons with experiments. A more detailed analysis (experiments versus theory) is presented in [25, 24]. A more theoretical approach, with the focus on genetic heterogeneity, can be found in [28]. More recent papers look at what happens when individuals are able to cooperate [29, 33].

One fact that has not been mentioned before is that the Voter model [6], which is closely related to the Moran process, can be mapped onto a process in which random walkers on a lattice coalesce (annihilate) when they meet. The similarity between coalescing random walk and annihilating sector boundary walls discussed above should now be obvious. Reference [7] is a nice review article discussing connections between various evolutionary models, coalescing random walks, and more general coalescence processes.

Exercise 5 (easy)

Consider the model from Section 3.2. Show that the number of different, neutral genotypes present at the front scales with its vertical position y as $\sim n_0/\sqrt{D(y-y_0)}$, where n_0 is the number of genotypes at y_0. Hint: how many sectors of size $\langle w(y)\rangle$ can one fit into a fixed length L?

Exercise 6 (hard)

In the case of a flat, vertically expanding front from Section 3.2, only one genotype will eventually survive genetic drift and reach fixation at the front. Coexistence of different genotypes is not possible in this model. However, things change if the population expands radially as in Fig. 10, left.

1) Show that in this case, a neutral mutant B can either get extinct or reach fixation at the front but without wiping out the initial genotype A, thus both genotypes can coexist for $r \to \infty$. Assume the mutant occupies a sector of angular size ϕ_0 when the colony radius is r_0.

Hint: show that, conditioned on fixation, the angular size ϕ of sector B at radius r will change with r according to the following equation (assuming $\phi \ll 2\pi$):

$$\partial_r P(\phi, r|\phi_0, r_0) = \frac{2D}{r^2}\partial_\phi^2 P(\phi, r|\phi_0, r_0), \tag{93}$$

with boundary conditions $P(0, r|\phi_0, r_0) = P(\infty, r|\phi_0, r_0) = 0$ where $P(\phi, r|\phi_0, r_0)$ is the probability that the sector has size ϕ at r if its size was ϕ_0 at r_0. Solve this equation using the method of images:

$$P(\phi, r|\phi_0, r_0) = G(\phi, r|\phi_0, r_0) - G(-\phi, r|\phi_0, r_0), \tag{94}$$

where $G(\phi, r|\phi_0, r_0)$ is the solution of Eq. (93) without annihilation (that is, for boundary conditions $P(-\infty, r|\phi_0, r_0) = P(\infty, r|\phi_0, r_0) = 0$). Show that

$$G(\phi, r|\phi_0, r_0) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\phi - \phi_0)^2}{2\sigma^2}\right), \tag{95}$$

where $\sigma^2 = \sigma^2(r, r_0) = 4D\int_{r_0}^{r} 1/r'^2 dr' = 4D(r_0^{-1} - r^{-1})$. Finally, calculate the mean angular sector size $\langle\phi(r \to \infty)\rangle$.

2) How many different, neutral genotypes will be present at the front if their number is n_0 at r_0? Hint: how many sectors of size $\langle\phi(r \to \infty)\rangle$ can one fit into the (angular) perimeter of a circle?

References

[1] M. Ancliff and J-M. Park. Spin coherent state representation of the crow-kimura and eigen models of quasispecies theory. *Journal of Statistical Physics*, 143(4):636–656, May 2011.

[2] T. Antal and P. L. Krapivsky. Exact solution of a two-type branching process: models of tumor progression. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(08):P08018, August 2011.

[3] T. Antal, P. L. Krapivsky, and M. A. Nowak. Spatial evolution of tumors with successive driver mutations. *arXiv preprint arXiv:1308.1564*, 2013.

[4] E. Baake, M. Baake, and H. Wagner. Ising quantum chain is equivalent to a model of biological evolution. *Physical review letters*, 78(3):559–562, 1997.

[5] E. Baake, H. Wagner, et al. Mutation-selection models solved exactly with methods of statistical mechanics. *Genetical research*, 78(1):93–117, 2001.

[6] E. Ben-Naim, L. Frachebourg, and P. L. Krapivsky. Coarsening and persistence in the voter model. *Physical Review E*, 53(4):3078, 1996.

[7] N. Berestycki. Recent progress in coalescent theory. *Ensaios Matematicos*, 16:1–193, 2009.

[8] M. Broom and J. Rychtář. An analysis of the fixation probability of a mutant on special classes of non-directed graphs. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 464(2098):2609–2627, October 2008.

[9] H. P. de Vladar and N. H. Barton. The contribution of statistical physics to evolutionary biology. *Trends in ecology & evolution*, 2011.

[10] L. Demetrius. Random spin models and chemical kinetics. *The Journal of chemical physics*, 87:6939, 1987.

[11] W. Ebeling, A. Engel, B. Esser, and R. Feistel. Diffusion and reaction in random media and models of evolution processes. *Journal of Statistical Physics*, 37(3):369–384, 1984.

[12] M. Eden. A two-dimensional growth process. *Dynamics of fractal surfaces*, page 265–283, 1961.

[13] M. Eigen and P. Schuster. A principle of natural self-organization. *Naturwissenschaften*, 64(11):541–565, 1977.

[14] C. L. Epstein. Anderson localization, non-linearity and stable genetic diversity. *Journal of Statistical Physics*, 124(1):25–46, July 2006.

[15] W. J. Ewens. *Mathematical Population Genetics. I. Theoretical Introduction.* Springer, 2004.

[16] F. D. C. Farrell, O. Hallatschek, D. Marenduzzo, and B. Waclaw. Mechanically driven growth of quasi-two-dimensional microbial colonies. *Physical Review Letters*, 111(16), October 2013.

[17] S. Galluccio. Exact solution of the quasispecies model in a sharply peaked fitness landscape. *Physical Review E*, 56(4):4526, 1997.

[18] C. Godrèche and J. M. Luck. Dynamics of the condensate in zero-range processes. *Journal of Physics A: Mathematical and General*, 38(33):7215, August 2005.

[19] Matti Gralka, Fabian Stiewe, Fred Farrell, Wolfram Möbius, Bartlomiej Waclaw, and Oskar Hallatschek. Allele surfing promotes microbial adaptation from standing variation. *Ecology Letters*, page 12625, June 2016.

[20] O. Hallatschek, P. Hersen, S. Ramanathan, and D. R. Nelson. Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences*, 104(50):19926, 2007.

[21] O. Hallatschek and D. R. Nelson. Life at the front of an expanding population. *Evolution*, 64(1):193–206, 2010.

[22] K. Jain and J. Krug. Deterministic and stochastic regimes of asexual evolution on rugged fitness landscapes. *Genetics*, 175(3):1275–1288, December 2006.

[23] Peter Keller and Tibor Antal. Mutant number distribution in an exponentially growing population. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(1):P01011, January 2015.

[24] K. S. Korolev, M. J. I. Müller, N. Karahan, A. W. Murray, O. Hallatschek, and D. R. Nelson. Selective sweeps in growing microbial colonies. *Physical Biology*, 9(2):026008, April 2012.

[25] K.S. Korolev, J.B. Xavier, D.R. Nelson, and K.R. Foster. Data from: A quantitative test of population genetics using spatio-genetic patterns in bacterial colonies. 2011.

[26] J. Krug and C. Karl. Punctuated evolution for the quasispecies model. *Physica A: Statistical Mechanics and its Applications*, 318(1):137–143, 2003.

[27] E. Kussell, S. Leibler, and A. Grosberg. Polymer-population mapping and localization in the space of phenotypes. *Physical Review Letters*, 97(6), August 2006.

[28] M. O. Lavrentovich, K. S. Korolev, and D. R. Nelson. Radial domany-kinzel models with mutation and selection. *arXiv preprint arXiv:1209.4103*, 2012.

[29] M. O. Lavrentovich and D. R. Nelson. Asymmetric mutualism in two-and three-dimensional range expansions. *Physical review letters*, 112(13):138102, 2014.

[30] I. Leuthäusser. Statistical mechanics of eigen's evolution model. *Journal of statistical physics*, 48(1):343–360, 1987.

[31] E. Lieberman, Ch. Hauert, and M. A. Nowak. Evolutionary dynamics on graphs. *Nature*, 433(7023):312–316, January 2005.

[32] H. J. Muller. Some genetic aspects of sex. *The American Naturalist*, 66(703):118–138, March 1932.

[33] M. J. I. Müller, B. I. Neugeboren, D. R. Nelson, and A. W. Murray. Genetic drift opposes mutualism during spatial population expansion. *Proceedings of the National Academy of Sciences*, 111(3):1037–1042, 2014.

[34] K. P. N. Murthy and K. W. Kehr. Mean first-passage time of random walks on a random lattice. *Physical Review A*, 40(4):2082–2087, August 1989.

[35] M. A. Nowak. What is quasispecies? *Trends in Ecology and Evolution*, 7:118, 1992.

[36] M. A. Nowak. *Evolutionary Dynamics*. Harvard University Press, September 2006.

[37] J.-M. Park and M. W. Deem. Phase diagrams of quasispecies theory with recombination and horizontal gene transfer. *Physical Review Letters*, 98(5), January 2007.

[38] A.Z. Rosenthal and M.B. Elowitz. Following evolution of bacterial antibiotic resistance in real time. *Nature Genetics*, 44(1):11–13, 2011.

[39] Hal L. Smith and Paul E. Waltman. *The Theory of the Chemostat: Dynamics of Microbial Competition*. Cambridge University Press, January 1995.

[40] B. Waclaw, R. J. Allen, and M. R. Evans. Dynamical phase transition in a model for evolution with migration. *Physical Review Letters*, 105(26), 2010.

[41] C. O. Wilke. Quasispecies theory in the context of population genetics. *BMC evolutionary biology*, 5(1):44, 2005.

[42] G. Woodcock and P. G. Higgs. Population evolution on a multiplicative single-peak fitness landscape. *J. Theor. Biol.*, 179:61, 1996.

The physics of biological evolution - exercises

Bartek Waclaw

School of Physics and Astronomy, The University of Edinburgh

July 2018

Exercise 1 (easy)

Assume we have only one species A which reproduces with rate b (probability per unit time) and dies with rate d, that is

$$A \to 2A \quad \text{with rate b,} \tag{1}$$
$$A \to 0 \quad \text{with rate d.} \tag{2}$$

Let there be only 1 A at $t = 0$ and $b > d$. Show that (1) the average number of A increases in time as $\sim \exp((b-d)t)$, (2) the probability that A goes extinct (i.e. there remains no A after some time) is d/b.

Exercise 2 (medium)

Take the Moran model where B has fitness $r = 1 + s$ and A has fitness 1, and the population size is N. Physical time is defined as before: $T = n/N$ where n is the number of steps. There is only one B at $T = 0$.

1) How much time does it take (to the leading order) for species B to eliminate A from the population (i.e. fixation of B) if $s = r - 1 \gg 1/N$ and N is large?

2) How much time does that take (again, to the leading order) if $s < 0$?

3) Assume the population initially consists of A only. When A reproduces, it can mutate into B with some small probability $\gamma \ll 1$. If B has a selective advantage $s > 0$ and $\gamma N \ln N \ll s$, what is the average time it takes for B to reach fixation?

Exercise 3 (easy)

Take the quasispecies model from Lecture 2 and assume the same fitness for all genotypes. How does the ground state of such "neutral" quasispecies model ($R(z) = 1$) look like? What is the average magnetization z?

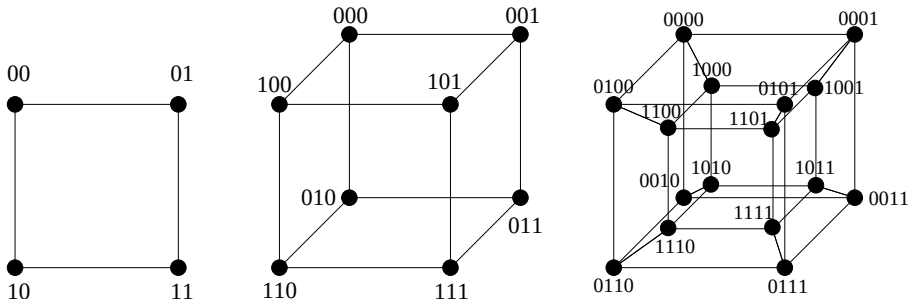


Figure 1: Examples of mutation graphs for the para-mu-se model, for $L = 2, 3, 4$.

Exercise 4 (medium)

Consider the following alternative to the Eigen and Schuster quasispecies model:

$$\frac{\mathrm{d}}{\mathrm{d}t} n_i = n_i(r_i - J(t)) + \sum_{j=0}^{2^L-1} Q_{ij} n_j, \tag{3}$$

where the mutation matrix reads

$$Q_{ij} = \begin{cases} \gamma & \text{if } d(i,j) = 1 \\ -L\gamma & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

and γ is now the mutation rate, not probability. In this model, mutations are allowed only along the edges of a hypercubic graph of dimension L, see Fig. 1. This model is called the para-mu-se model (parallel mutation and selection). Calculate the error threshold for this model, assuming (as in the notes) that $r_0 = A > 1$ and $r_1 = \cdots = r_{2^L-1} = 1$.

Exercise 5 (easy)

Consider the model from Section 3.2. Show that the number of different, neutral genotypes present at the front scales with its vertical position y as $\sim n_0/\sqrt{D(y - y_0)}$, where n_0 is the number of genotypes at y_0. Hint: how many sectors of size $\langle w(y) \rangle$ can one fit into a fixed length L?

Exercise 6 (hard)

In the case of a flat, vertically expanding front from Section 3.2, only one genotype will eventually survive genetic drift and reach fixation at the front. Coexistence of different genotypes is not possible in this model. However, things change if the population expands radially as in Fig. 10, left, in the notes.

1) Show that in this case, a neutral mutant B can either get extinct or reach fixation at the front but without wiping out the initial genotype A, thus both genotypes can coexist for $r \to \infty$.

Assume the mutant occupies a sector of angular size ϕ_0 when the colony radius is r_0.

Hint: show that, conditioned on fixation, the angular size ϕ of sector B at radius r will change with r according to the following equation (assuming $\phi \ll 2\pi$):

$$\partial_r P(\phi, r|\phi_0, r_0) = \frac{2D}{r^2}\partial_\phi^2 P(\phi, r|\phi_0, r_0), \tag{5}$$

with boundary conditions $P(0, r|\phi_0, r_0) = P(\infty, r|\phi_0, r_0) = 0$ where $P(\phi, r|\phi_0, r_0)$ is the probability that the sector has size ϕ at r if its size was ϕ_0 at r_0. Solve this equation using the method of images:

$$P(\phi, r|\phi_0, r_0) = G(\phi, r|\phi_0, r_0) - G(-\phi, r|\phi_0, r_0), \tag{6}$$

where $G(\phi, r|\phi_0, r_0)$ is the solution of Eq. (5) without annihilation (that is, for boundary conditions $P(-\infty, r|\phi_0, r_0) = P(\infty, r|\phi_0, r_0) = 0$). Show that

$$G(\phi, r|\phi_0, r_0) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\phi - \phi_0)^2}{2\sigma^2}\right), \tag{7}$$

where $\sigma^2 = \sigma^2(r, r_0) = 4D\int_{r_0}^{r} 1/r'^2 dr' = 4D(r_0^{-1} - r^{-1})$. Finally, calculate the mean angular sector size $\langle\phi(r \to \infty)\rangle$.

2) How many different, neutral genotypes will be present at the front if their number was n_0 at r_0? Hint: how many sectors of size $\langle\phi(r \to \infty)\rangle$ can one fit into the (angular) perimeter of a circle?

CQF

Superfluids & Superconductors

Lecture notes by Derek Lee

Imperial College London

Physics by the Lake 2018

Contents

1 Neutral Superfluids **3**

1 Ideal Bose gas and Bose condensation . . . 3
1.1 Ground state of an ideal Bose gas . . . 3
1.2 Bose condensation at non-zero temperature . . . 4
1.3 Measuring the momentum distribution . . . 5
1.4 ...but the ideal Bose gas is pathological . . . 6
2 What is superfluidity? . . . 7
2.1 Landau criterion for superfluidity . . . 7
3 Measuring superfluidity . . . 9
4 Interacting Bose liquid . . . 10
4.1 Condensate wavefunction . . . 10
4.2 Nonlinear Schrödinger equation . . . 12
4.3 Healing length . . . 13
5 Vortices . . . 13
5.1 Circulation quantisation . . . 14
5.2 Vortex line energy . . . 15
5.3 Rotating superfluids . . . 16
6 Dynamics . . . 17
6.1 Gross-Pitaevskii equation . . . 17
6.2 Hydrodynamics . . . 18
6.3 Sound waves . . . 19
7 Superfluidity, spontaneous symmetry breaking and Goldstone modes . . . 20

2 Charged superfluids **23**

1 Overview . . . 23
2 Phenomenological theory . . . 25
2.1 Ginzburg-Landau theory . . . 26
3 Magnetic Response . . . 27
3.1 London equation and Meissner effect . . . 27
3.2 Perfect conductor *vs* superconductor . . . 29
4 Flux quantisation . . . 30
4.1 Flux quantum . . . 30
4.2 Flux lines . . . 32
4.3 Two types of superconductors . . . 32
4.4 Comparison with neutral superfluids . . . 33
5 Josephson Effect . . . 34
5.1 DC Josephson current . . . 34

5.2 Interferometry for Cooper pairs 36
5.3 AC Josephson Effect 38

3 Cooper Pairs **39**
1 Effective electron attraction 40
2 Cooper Instability 42
3 BCS Ground State 46
3.1 Variational Ground State 46
3.2 Variational approach 48
3.3 Condensation energy 50
3.4 Isotope effect 51
4 Gap function & pairing symmetry 51
5 BCS Excitations 53
5.1 Excitations of a Fermi gas 53
5.2 Excitations for the BCS state 55
5.3 Superconductivity 57
6 Summary 58

Superfluids & Superconductors
Chapter 1

Neutral Superfluids

- Ideal Bose gas
  - Ground state is the totally condensed state
  - Heuristic argument for the critical temperature for Bose condensation
- What is superfluidity?
  - Landau criterion for superfluidity
  - ideal Bose gas is not a superfluid
- Interacting Bose liquid
  - characteristic length: healing length
- Vortex lines
  - circulation quantised in units of h/m: macroscopic manifestation of angular momentum quantisation
  - rotating superfluids: lower and upper critical angular frequencies

Bosons exhibit superfluidity at low temperatures. Superfluidity is linked to, but not synonymous with, Bose-Einstein condensation. This is the phenomenon where a macroscopic number of bosons occupy the same state.

1 Ideal Bose gas and Bose condensation

We start with an introduction to Bose condensation using the simple example of the *ideal Bose gas* where the bosons do not interact with each other. This is obvious at zero temperature: all the non-interacting bosons occupy the same state independently. In fact, Bose condensation persists in the 3D ideal Bose gas until a certain critical temperature T_c is reached. In the next chapter, we will see that Bose condensation also survives weak interactions, even though the ground state of an interacting Bose liquid cannot be simply described by the occupation of individual single-particle eigenstates.

1.1 Ground state of an ideal Bose gas

Consider N non-interacting bosons of mass m with zero spin in a potential $V(\mathbf{r})$. The ground state for a single particle in this potential well has the wavefunction $u_0(\mathbf{r})$ with energy ϵ_0. For a homogenous system, u_0 is the zero-momentum state ($\mathbf{p} = 0$). In a simple harmonic well with natural frequency ω_0, this is a Gaussian wavepacket with radial size $(\hbar/m\omega_0)^{1/2}$. In general, unless there are special symmetry reasons, this ground state is unique.

For N bosons, we have to discuss a *many-body wavefunction* $\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ which is the probability amplitude of finding the N particles at positions $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N$. Remember that these are indistinguishable bosons, so we cannot say that which particle is at $\mathbf{r}_1$. Mathematically, the exchange of any of the particle coordinates does not change the wavefunction: *e.g.* $\Psi(\ldots, \mathbf{r}_i, \ldots, \mathbf{r}_j, \ldots) = \Psi(\ldots, \mathbf{r}_j, \ldots, \mathbf{r}_i, \ldots)$.

The ground state of this ideal Bose gas is totally condensed: all the bosons occupy the single-particle eigenstate u_0:

$$\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \prod_{i=1}^{N} u_0(\mathbf{r}_i) \tag{1.1}$$

1.2 Bose condensation at non-zero temperature

Does Bose condensation survive at non-zero temperatures? As we warm up the system, bosons will be thermally excited to states of higher energy so that the ideal Bose gas is no longer totally condensed in the u_0 state (Fig. 1.1). Nevertheless, Bose condensation can exist at low temperatures in the sense that there is a **macroscopic occupation** of the u_0 state: the number of particles N_0 in this state, although no longer equal to the total number of particles N, is still proportional to the volume of the system. In other words, a finite fraction of the particles occupy a single state.

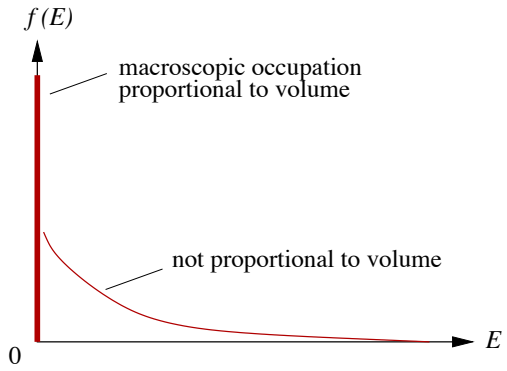


Figure 1.1: Occupation $f(E)$ of states with energy E: macroscopic occupation of the lowest-energy state, plus non-macroscopic occupation of higher states.

This macroscopic occupation of a single-particle state persists up to a critical temperature T_c (Fig. 1.2). This temperature scale is controlled by quantum mechanics, as can be seen in the following heuristic argument for a homogeneous system.

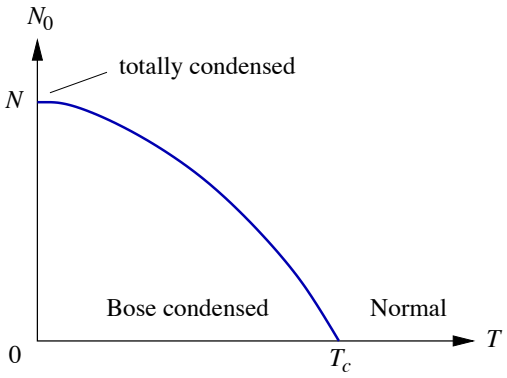


Figure 1.2: Bose-Einstein condensation: Number of particles N_0 condensed into the $\mathbf{p} = 0$ state as a function of temperature T. The condensate fraction N_0/N reduces with increasing temperature as $1 - (T/T_c)^{3/2}$ and vanishes above T_c.

At temperature T, particles are excited to states with energy up to the thermal energy $k_{\rm B}T$. The typical wavevector k_T for such states will therefore be given by $\hbar^2 k_T^2/2m \sim k_{\rm B}T$, so that $k_T \sim \sqrt{mk_{\rm B}T}/\hbar$ (ignoring factors of 2). The typical wavelength will be $\lambda_T = 2\pi/k_T$. This is just a numerical factor from the conventional **thermal deBroglie wavelength**:

$$\lambda_{\rm dB} = \sqrt{\frac{2\pi\hbar^2}{mk_{\rm B}T}} \tag{1.2}$$

Quantum effects become important when the thermal wavelength exceeds the interparticle spacing. For a gas with number density n, the interparticle spacing is $1/n^{1/D}$. So, quantum effects become significant when $n\lambda_{\rm dB}^D > 1$. The lower the density, the lower is the temperature at which quantum physics come into play. The crossover temperature is called the **degeneracy temperature** T_0. The inequality gives us an estimate of this degeneracy temperature:

$$k_{\rm B}T_0(n) = \frac{2\pi\hbar^2}{m}\, n^{2/D} \tag{1.3}$$

below which we expect quantum effects to become significant. We see that, if the ideal Bose gas is dilute, then we have to go to very low temperatures before we see quantum effects.

A quantitative calculation using the Bose-Einstein distribution for the thermal occupation numbers confirms that the above heuristic argument captures the physics of Bose condensation. In three dimensions, the critical temperature for Bose condensation in an infinite homogeneous system is only different from the degeneracy temperature by a numerical factor: $T_c = 0.527T_0$.

1.3 Measuring the momentum distribution

Ultracold atoms in an optical trap

Alkali atoms, such as ^{87}Rb, can be captured and laser-cooled in optical or magnetic traps down to nanokelvin temperatures. These atomic vapours are metastable states which exhibit Bose condensation (before they eventually coaslesce into a solid.) They are very dilute and so the particles only interact with each other weakly. So, the system can be considered as a nearly ideal Bose gas. The particles are condensed into the lowest single-particle eigenstate (a Gaussian wavepacket) of the harmonic well that traps them.

To measure the momentum distribution, the trap is switched off so that the tightly confined atomic cloud can now expand. This is driven by the quantum cost in kinetic energy for confinement. As the cloud expands to a size larger than optical wavelengths, it can be imaged with optical light. One can examine the shadow it casts and find out the velocity of the atoms simply from how far they have travelled from the original position of the trap. For a particle with mass m and momentum $\mathbf{p}$, this distance would be $\mathbf{R} = \mathbf{p}/mt = \mathbf{v}/t$ where t is the time that has elapsed since the trap was switched off. In other words, the density profile $\rho(\mathbf{R}, t)$ of this cloud as a function of radial distance from the original trap gives us the distribution for the momentum of the particles: $n(\mathbf{p}) = \rho(\mathbf{p}/mt, t)$.

Bose-Einstein condensation in ultracold atoms was first observed by Cornell and Wieman in Colorado in 1995 and independently by Ketterle at Massachusetts Institute of Technology. They found that the ultracold atomic gases are nearly totally Bose-condensed, as

we might expect. They were awarded the Nobel Prize in 2001. For more information, see `http://jilawww.colorado.edu/bec/` and `http://cua.mit.edu/ketterle_group/`.

Liquid helium

Liquid ^{4}He has much higher particle density than the atomic vapours. It is a strongly interacting Bose liquid and the fraction of helium atoms condensed in the $\mathbf{p} = 0$ state is only about 10%. Inelastic neutron scattering is used to measure the momentum distribution in this sytem. Fast neutrons hit the slow-moving helium atoms and scatter off them. From the energy and momentum of the scattered neutrons that emerge from the sample, we can work out the energy and momentum picked up by the helium atoms at the collisions. Simple kinematics show that this energy/momentum transfer ΔE depends on the initial momentum $\mathbf{p}$ of the helium atoms: $\Delta E = \mathbf{P}^2/2m + \mathbf{p}\cdot\mathbf{P}/m$ where $\mathbf{P}$ is the neutron momentum. Therefore, we can deduce the original momentum distribution of the particles from the scattering cross-section of the neutrons[1].

1.4 ...but the ideal Bose gas is pathological

The ideal Bose gas is unfortunately of limited use as a model of physical systems.

- Consider an ideal Bose gas in a macroscopic container with a micron-sized potential well deep enough to support a bound state. *All* the bosons will condense into this micron-sized bound state! Realistically, repulsion between the bosons would prevent this collapse so that the ground state would have more or less uniform density profile occupying the whole container.
- The ideal Bose gas does not have zero viscosity. We will see this when we discuss the Landau criterion for superfluidity in section (2.1).
- Repulsive bosons have different low-energy dynamics from the ideal Bose gas. We will later calculate the dispersion relation for the low-energy long-wavelength excitations of an interacting Bose liquid.

Due to these problems, we are driven to study the **interacting Bose liquid**. This should be contrasted with electron systems where a model of non-interacting electrons goes a long way in explaining the electronic properties of insulators, semiconductors and metals.

In the presence of interactions, the totally condensed state (1.1) is no longer an exact eigenstate of the system. Does the condensate survive to some extent? The answer is yes, even when the repulsive interactions between the particles are quite strong. In particular, liquid helium has a high enough density of particles that their mutual interactions cannot be ignored. (After all, they have condensed from a gas to a liquid already.) In fact, only about 10% of the helium atoms remains condensed in the zero-momentum state even at the lowest experimental temperatures ($T \ll T_c$). For the rest of this chapter, we will discuss the statics and dynamics of an interacting Bose liquid *assuming that the condensate has not been totally destroyed*.

[1] P.C. Hohenberg and P.M. Platzman, *Phys. Rev.* **152** 198 (1966).

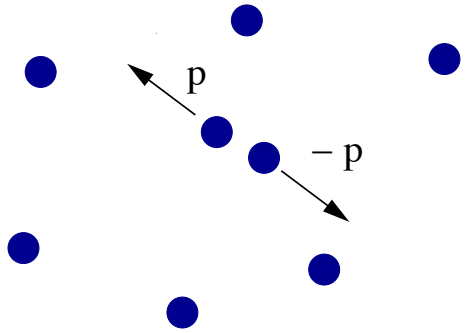


Figure 1.3: Condensate depletion. Collisions between two particles at zero momenta can knock them out of the condensate into states with opposite momenta: $|0,0\rangle \to |\mathbf{p},-\mathbf{p}\rangle$. These are zero-point processes which occur even at zero temperature.

2 What is superfluidity?

2.1 Landau criterion for superfluidity

Suppose a normal fluid is flowing inside a long straight pipe. The viscous friction between the fluid and the pipe walls slows down the fluid which eventually comes to rest relative to the pipe. Microscopically, the fluid loses momentum through the scattering of atoms against the pipe wall.

A superfluid has zero viscosity. It is somehow unable to lose its momentum through scattering with the pipe wall. What is preventing these processes? Landau realised that the absence of scattering does not depend on details of atom-atom interactions but something more fundamental: *conservation laws might prevent the scattering of individual particles from happening.* He used this idea to formulate a very general microscopic criterion for superfluidity in terms of the excitation energies of the system.

To be more specific, let us consider a fluid flowing in a pipe, at a velocity $-\mathbf{v}$ relative to the pipe. Let us denote the rest frame of the fluid by $Z_{\rm fluid}$, and the laboratory frame, where the pipe is at rest, by $Z_{\rm lab}$. In $Z_{\rm fluid}$, the stationary fluid should be in its (Bose-condensed) ground state, and the pipe is moving at $+\mathbf{v}$. To distinguish between quantities measured in the two frames, quantities measured in the lab frame will be labelled by a prime, *e.g.* $\mathbf{r}'$ for positions. Let us consider elastic scattering of the bosons by the pipe[2]. In the lab

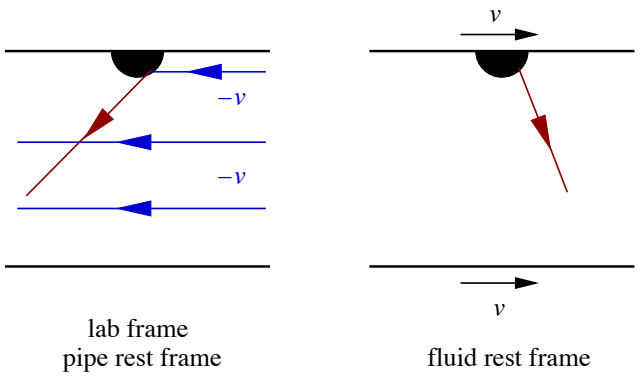


Figure 1.4: Fluid flowing through a pipe, viewed from the pipe rest frame $Z_{\rm lab}$ and the fluid rest frame $Z_{\rm fluid}$.

frame $Z_{\rm lab}$, the pipe wall presents a stationary potential $V(\mathbf{r})$ to the moving atoms. This

[2]The argument can be easily extended to inelastic scattering.

can be treated as a perturbation to the Hamiltonian:

$$\delta\hat{H} = \int V(\mathbf{r})\hat{n}(\mathbf{r})\, d^3\mathbf{r} = \sum_{\mathbf{kq}} V_{\mathbf{q}} \hat{a}^\dagger_{\mathbf{k}-\mathbf{q}} \hat{a}_{\mathbf{k}} \tag{1.4}$$

where $V_{\mathbf{q}}$ is the Fourier transform of the potential and $\hat{a}^\dagger_{\mathbf{k}}$ creates a boson with momentum $\hbar\mathbf{k}$. We see that this potential scatters individual atoms, randomising the direction of motion of the atoms so that fluid loses its overall momentum.

Landau pointed out that these scattering processes are prevented if the possible final states for the bosons are all higher in energy than the initial state:

$$\Delta E' = E'_{\text{final}} - E'_{\text{init}} > 0 \tag{1.5}$$

It is simplest to express this criterion in terms of the excitations of the fluid in its rest frame Z_{fluid} where the initial state is the ground state of a stationary $\mathbf{p} = 0$ condensate. To transform dynamical quantities between non-relativistic inertial frames, we need to use the **Galilean transformation**. A boson of mass m with momentum of $\mathbf{p}$ in Z_{fluid} would appear to have momentum $\mathbf{p}' = \mathbf{p} - m\mathbf{v}$ in the pipe frame Z_{lab}. So, the total momentum of the system $\mathbf{P} = \sum_i \mathbf{p}_i$ transforms as $\mathbf{P}' = \mathbf{P} - Nm\mathbf{v}$. The energy

$$E = \sum_i \frac{\mathbf{p}_i^2}{2m} + \sum_{i>j} U(\mathbf{r}_i - \mathbf{r}_j) \tag{1.6}$$

in the fluid rest frame Z_{fluid} becomes:

$$E' = \sum_i \frac{(\mathbf{p}_i - m\mathbf{v})^2}{2m} + \sum_{i>j} U(\mathbf{r}'_i - \mathbf{r}'_j) = E - \mathbf{v}\cdot\mathbf{P} + \frac{1}{2}Nm\mathbf{v}^2 \tag{1.7}$$

in the pipe rest frame Z_{lab}. Suppose the system changed from an initial state to a final state, and the energy changed by $\Delta E = E_{\text{final}} - E_{\text{init}}$ and the momentum changed by $\Delta\mathbf{P} = \mathbf{P}_{\text{final}} - \mathbf{P}_{\text{init}}$ as measured in the fluid rest frame Z_{fluid}. In the pipe frame Z_{lab}, the relative energy and momentum are related to the fluid-frame quantities by:

$$\Delta E' = \Delta E - \mathbf{v}\cdot\Delta\mathbf{P}\,, \qquad \Delta\mathbf{P}' = \Delta\mathbf{P} \tag{1.8}$$

We now see that the no-scattering condition (1.5) of $\Delta E' > 0$ becomes a condition on the energy and momentum of the excitations of the fluid at rest, *i.e.* the excitations from the true ground state of this system. This **Landau criterion** states that the system is a superfluid if the energies ΔE and momenta $\Delta\mathbf{P}$ of all the excitations accessible by single-particle scattering satisfy the inequality:

$$\boxed{\Delta E - \mathbf{v}\cdot\Delta\mathbf{P} > 0 \qquad \text{for all excitations}} \tag{1.9}$$

for a small but non-zero fluid velocity $\mathbf{v}$.

Let us apply the Landau criterion to the ideal Bose gas. The single-particle excitations simply taking a particle out of the condensate and putting it in a state with momentum $\Delta\mathbf{P} = \mathbf{p}$. This costs an energy of $\Delta E = \mathbf{p}^2/2m$. So, $\Delta E' = \Delta E - \mathbf{v}\cdot\Delta\mathbf{P} = \mathbf{p}^2/2m - \mathbf{v}\cdot\mathbf{p}$. This is sketched on the left in Fig. 1.5. We see that we can always find a suitable $\mathbf{p}$ at small momentum for the scattering process to be allowed ($\Delta E' = 0$ for elastic scattering, $\Delta E' < 0$ for inelastic scattering). Therefore, we come to the conclusion that the ideal

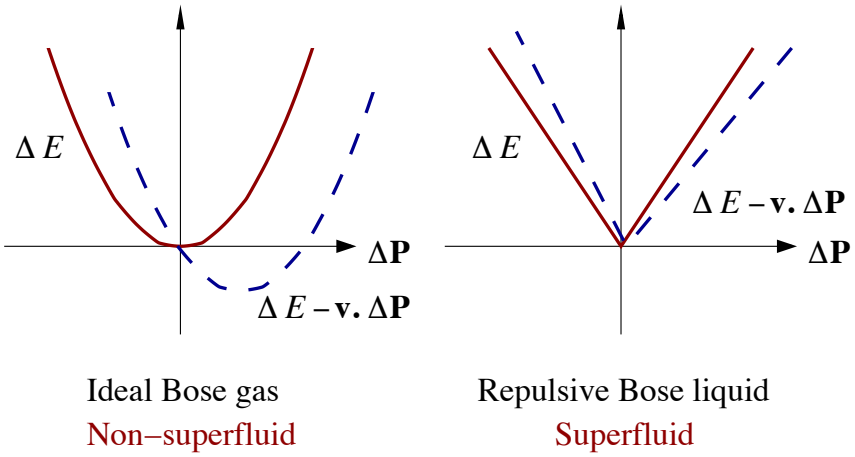


Figure 1.5: Galilean transformation for the excitation spectrum. Superfluidity requires that excitation energies remain positive after the transformation to the pipe rest frame: $\Delta E \to \Delta E - \mathbf{v} \cdot \Delta \mathbf{P}$.

Bose condensate will have a viscous drag with the walls of the pipe. **The ideal Bose gas is not a superfluid** even though it can be totally Bose condensed at zero temperature.

When can the Landau criterion be satisfied? For a small fluid velocity $\mathbf{v}$, this happens if the minimum phase velocity of the spectrum is non-zero. For instance, *sound waves* have the spectrum $\Delta E_{\mathbf{p}} = v_s|\mathbf{p}|$. So, the excitation energies in the pipe frame remain positive as long as $|\mathbf{v}| < v_s$. (See right diagram in Fig. 1.5.) In fact, it can be shown that this sound mode is characteristic of an interacting Bose liquid and is the only mode at long wavelengths. (Check the excitation spectrum for liquid helium ^{4}He in Fig. 2.1.) This allows us to conclude that an interacting Bose liquid is a superfluid.

Notice that the Landau criterion cannot be satisfied for a sufficiently large $\mathbf{v}$. Therefore, superfluidity is destroyed above a certain **critical velocity**. In this example, the critical velocity is the sound velocity v_s. In general, we may have missed out other mechanisms for dissipating momentum, so v_s is an upper bound for the critical velocity.

3 Measuring superfluidity

To measure the superfluidity of liquid helium directly, we have to measure the fraction of the liquid that is viscous. Andronikashvili was the first to devise such an experiment[3]. A stack of thin, closely spaced circular plates are suspended on a thin wire. The plates are immersed in a cylinder of liquid helium. The torque on the plates is proportional to the angle ϕ through which the wire has been twisted. Therefore the torsional oscillations follow simple harmonic motion in the angle of oscillations with period $P = 2\pi(I/G)^{1/2}$ where I is the moment of inertial of the rotating system, and G is the torque per unit radian of twist. The moment of inertia I includes that of the stack of disks and the normal viscous fluid that rotates with it: $I = I_{\text{plate}} + I_{\text{normal}}$. Since the plates are typically much heavier than the liquid, it is easy to see that any fractional change in the torsional period $\Delta P/P$ is related to the fractional change in the moment of inertia by: $\Delta P/P \simeq \Delta I/2I \ll 1$.

[3]E.L. Andronikashvili, *J. Phys. USSR* **10** 201 (1946).

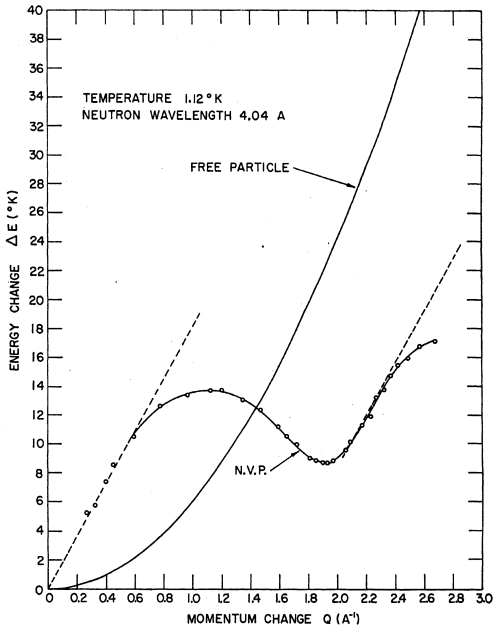


Figure 1.6: Excitation spectrum for liquid ^{4}He. The vertical axis is the excitation energy converted to a temperature, *i.e.* $\Delta E/k_{\rm B}$. The horizontal axis gives the wavevector $k = p/\hbar$ in inverse Angstroms. [D.G. Henshaw and A.D.B. Woods *Phys. Rev.* **121** 1266 (1961)]

The superfluid part of the fluid is not dragged by the disks and do not rotate with the plates. Therefore it does not contribute to the moment of inertia. So, when the system is cooled below the superfluid transition temperature T_c, the moment of inertial decreases and this is observed through a reduction of ΔP as the temperature is lowered. The experimental challenge is to measure these tiny fractional changes accurately.

In a recent experiment which claims to observe superflow in solid(!) helium[4], period shifts of 0 to 40ns out of a period of 1μs had to be detected accurately. They concluded that solid helium has a superfluid fraction of at most 1.7%. The origin of this superfluid fraction is not clear. Candidate mechanisms include the superfluidity of mobile vacancies in the solid and the condensation of rotons.

A similar experiment cannot be performed for trapped ultracold atoms. Instead, one can 'stir' the atomic clouds with laser light and observe quantised vortices. This is a signature of superfluidity that we will discuss in section 5.

4 Interacting Bose liquid

4.1 Condensate wavefunction

First of all, let us see how to extract physical information, such as the density and the current, about a totally condensed state (1.1). For a homogeneous system of volume V with periodic boundaries, the ground state is $u_0 = 1/\sqrt{V}$.

It is conventional to define a **condensate wavefunction** $\psi_{\rm c}(\mathbf{r})$ so that it is proportional to the macroscopically occupied wavefunction $u_0(\mathbf{r})$ but normalised so that $|\psi_{\rm c}(\mathbf{r})|^2$ gives

[4]E. Kim and M.H.W. Chan, *Science* **305** 1941 (2004); commentry by Tony Leggett on page 1921 of the same issue.

the local density of condensed particles:

$$\int |\psi_{\rm c}(\mathbf{r})|^2 d^3\mathbf{r} = N_0 \tag{1.10}$$

For the special case of the totally condensed state, this is the same as the total number density $N = N_0$ and the condensate wavefunction is simply $\psi_{\rm c} = \sqrt{N} u_0(\mathbf{r})$. More generally, the condensate wavefunction is a complex function that can vary spatially:

$$\psi_{\rm c}(\mathbf{r}) = n_c(\mathbf{r})^{1/2}\, e^{i\theta(\mathbf{r})} \tag{1.11}$$

Note that the wavefunction can be multiplied by a global constant phase factor $e^{i\theta_0}$ with no physical consequence. However, as we will see, a spatially varying phase does have physical meaning.

It is easy to see that the **boson density** $n(\mathbf{r})$ is simply $\langle \hat{n}(\mathbf{r})\rangle = \sum_{i=1}^{N} \langle \delta(\mathbf{r} - \mathbf{r}_i)\rangle = N|u(\mathbf{r})|^2 = |\psi_{\rm c}(\mathbf{r})|^2$.

We can also calculate the **current density J**, the rate at which particles flow through a cross-section per unit area of that cross-section. In a coarse-grained viewpoint of the fluid, the current density is related to the local fluid velocity $\mathbf{v}_s(\mathbf{r})$ by: $\mathbf{J}(\mathbf{r}) = n(\mathbf{r})\mathbf{v}_s(\mathbf{r})$. Microscopically, the current density operator should sum up the individual velocity contributions $\delta(\mathbf{r}-\mathbf{r}_i)(-i\hbar\boldsymbol{\nabla}_i/m)$ from each particle i. However, this operator is not Hermitian. So we need a symmetrised form to ensure that it can represent an observable[5]:

$$\hat{\mathbf{J}} = \frac{-i\hbar}{2m} \sum_i \left[\delta(\mathbf{r} - \mathbf{r}_i)\boldsymbol{\nabla}_i + \boldsymbol{\nabla}_i \delta(\mathbf{r} - \mathbf{r}_i)\right] \tag{1.12}$$

The local current density of a totally condensed state is:

$$\langle \hat{\mathbf{J}}(\mathbf{r})\rangle = \frac{-i\hbar}{2m}\left[\psi_{\rm c}^*(\mathbf{r})\boldsymbol{\nabla}\psi_{\rm c}(\mathbf{r}) - \psi_{\rm c}(\mathbf{r})\boldsymbol{\nabla}\psi_{\rm c}^*(\mathbf{r}) = \frac{\hbar}{m}\, n(\mathbf{r})\boldsymbol{\nabla}\theta(\mathbf{r})\right] \tag{1.13}$$

Note that the terms involving the gradients of $\sqrt{n}(\mathbf{r})$ cancel to give the simple relationship that **the gradient of the phase of the condensate wavefunction is proportional to the velocity of a moving condensate**. This leads us to an important expression for superfluid velocity $\mathbf{v}_s = \mathbf{J}/n$.

$$\mathbf{v}_s = \frac{\hbar}{m}\boldsymbol{\nabla}\theta \tag{1.14}$$

We have derived this for the totally condensed state. However, this relation between the condensate phase and the fluid velocity holds in general for any neutral Bose superfluid.

We can calculate the energy of the condensed state in the presence of interactions. In general, the interactions are pairwise with potential energy $U(\mathbf{r}_i - \mathbf{r}_j)$. To keep things general, suppose also that the Bose fluid also sees an external potential $V(\mathbf{r})$. The Hamiltonian is given by:

$$\hat{H} = \sum_{i=1}^{N}\left[-\frac{\hbar^2}{2m}\nabla_i^2 + V(\mathbf{r}_i)\right] + \sum_{i<j} U(\mathbf{r}_i - \mathbf{r}_j) \tag{1.15}$$

[5]The product of two Hermitian operators $\hat{A}$ and $\hat{B}$ that do not commute with each other is not Hermitian: $(\hat{A}\hat{B})^\dagger = \hat{B}\hat{A} \neq \hat{A}\hat{B}$. The only Hermitian linear combination is proportional to $\hat{A}\hat{B} + \hat{B}\hat{A}$.

Bose physics is often concerns low-energy states with wavelengths long compared to the interaction range of the atoms. So, we can approximate the pairwise interaction as a contact interaction[6]:

$$U(\mathbf{r}) = u\delta(\mathbf{r}) \tag{1.16}$$

Then, the energy of the totally condensed state is:

$$\langle \hat{H} \rangle = \int \left[\frac{\hbar^2}{2m} \int |\boldsymbol{\nabla}\psi_{\rm c}|^2 + V(\mathbf{r})|\psi_{\rm c}(\mathbf{r})|^2 + \frac{u}{2}|\psi_{\rm c}(\mathbf{r})|^4 \right] d^3\mathbf{r} \tag{1.17}$$

Note that there is a factor of $1/2$ in the interaction term to avoid double counting. The interaction term can be interpreted as each particle seeing an *averaged* interaction potential $V_{\rm H}(\mathbf{r})$ from the other particles. This is called the 'Hartree potential':

$$V_{\rm H}(\mathbf{r}) = \int d^3\mathbf{r}\, U(\mathbf{r}-\mathbf{r}')\langle \hat{n}(\mathbf{r}')\rangle = u|\psi_{\rm c}(\mathbf{r})|^2 \,. \tag{1.18}$$

4.2 Nonlinear Schrödinger equation

How does the condensate wavefunction behave in the presence of external potentials? We will adopt a *variational approach*. We *assume* a totally condensed state with wavefunction $\psi_{\rm c}$ and find the wavefunction which minimises the total energy of the system for a fixed number of bosons. Working in the grand canonical ensemble, we introduce a chemical potential μ and minimise $F[\psi_{\rm c}] = \langle \hat{H} - \mu\hat{N}\rangle$. From (1.10) and (1.17),

$$F = \int d^3\mathbf{r} \left[\frac{\hbar^2}{2m}|\boldsymbol{\nabla}\psi_{\rm c}|^2 + (V(\mathbf{r})-\mu)|\psi_{\rm c}|^2 + \frac{u}{2}|\psi_{\rm c}(\mathbf{r})|^4 \right] \tag{1.19}$$

To set the boundary conditions for this problem, we assume that the external potential is confined to a finite region so that the fluid should be uniform at infinity: $\psi_{\rm c}(\mathbf{r}\to\infty) = \sqrt{\bar{n}}$ for a system with average density $\bar{n}$. So, there is no kinetic energy to consider and the system should simply minimise $F_\infty = \int (u|\psi_{\rm c}|^4/2 - \mu|\psi_{\rm c}|^2) d^3\mathbf{r}$. This sets the chemical potential:

$$\mu = u|\psi_{\rm c}(\mathbf{r}\to\infty)|^2 = u\bar{n} \tag{1.20}$$

Now, we can use the calculus of variations to obtain the differential equation that the wavefunction obeys. To do so, we consider how F changes due to a small change in the wavefunction $\delta\psi_{\rm c}(\mathbf{r})$ with the condition that $\delta\psi_{\rm c}\to 0$ at infinity, and then we require that the change δF is zero to first order in $\delta\psi_{\rm c}$. It can be shown that: [**Exercise**: prove this.]

$$\delta F \;\simeq\; \int d^3\mathbf{r} \left[\delta\psi_{\rm c}^* \left(-\frac{\hbar^2}{2m}\nabla^2\psi_{\rm c} + V_{\rm eff}(\mathbf{r})\psi_{\rm c} \right) + \left(-\frac{\hbar^2}{2m}\nabla^2\psi_{\rm c}^* + V_{\rm eff}(\mathbf{r})\psi_{\rm c}^* \right) \delta\psi_{\rm c} \right] \tag{1.21}$$

where $V_{\rm eff} = V(\mathbf{r}) + V_{\rm H}(\mathbf{r}) - \mu$ can be regarded as an effective potential. In order that $\delta F = 0$, the condensate wavefunction has to satisfy a *non-linear* version of the Schrödinger equation for $\psi_{\rm c}$:

$$\boxed{ -\frac{\hbar^2}{2m}\nabla^2\psi_{\rm c} + (V + u|\psi_{\rm c}|^2 - \mu)\,\psi_{\rm c} = 0 \qquad \text{with} \quad \mu = u\bar{n} } \tag{1.22}$$

[6] A more rigorous treatment is to parametrise u so that the contact interaction reproduces the s-wave scattering length of the real interatomic potential.

4.3 Healing length

A simple application of the non-linear Schrödinger equation shows us that the interacting Bose system has a characteristic length scale, the 'healing length', which is absent from the ideal Bose gas.

Suppose we have a homogeneous semi-infinite system with a hard wall (infinite potential barrier) at $x = 0$. The system is translationally invariant in the y and z directions with periodic boundaries. So, the condensate wavefunction is constant in the y and z directions but has to vanish at the infinite barrier at $x = 0$. So, we have to solve the equation

$$\left[-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + u(|\psi_c|^2 - \bar{n})\right]\psi_c = 0 \tag{1.23}$$

An energy scale in this problem is the Hartree energy $u\bar{n}$. We know from the kinetic energy operator that $(\hbar^2/2m)(\text{length})^{-2}$ is also an energy. By comparing these two energies, we can extract a characteristic length

$$\xi = \frac{\hbar}{(mu\bar{n})^{1/2}} \tag{1.24}$$

In other words, we have extracted a length scale where the kinetic energy balances the interaction energy $u\bar{n}$. This is called the **healing length** or **coherence length** for reasons that will become clear below. For liquid helium and optically trapped ultracold atoms, it is a microscopic quantity of the order of the interparticle spacing.

Most non-linear second-order differential equations do not have an analytic solution. This one is fortunately an exception: $\psi_c(x) = \sqrt{\bar{n}}\tanh(x/\xi)$. [**Exercise**: check this.] We see that, although $\psi_c(x)$ is forced to vanish at $x = 0$, the wavefunction recovers to its bulk value of $\sqrt{\bar{n}}$ over a length scale of ξ (Fig. 1.7). This is why ξ is called the healing length in superfluid physics.

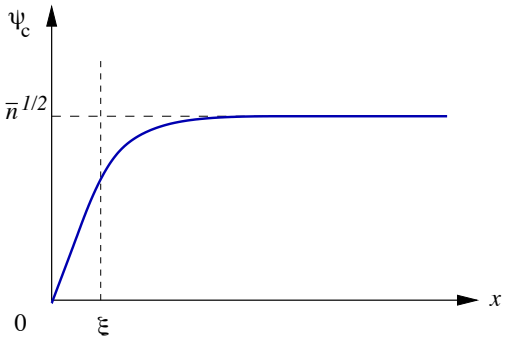


Figure 1.7: Condensate near a wall. The condensate wavefunction recovers to the bulk value over the scale of a healing length ξ.

5 Vortices

Superfluids support vortex lines. These lines of circulating superflow are topological defects in the condensate wavefunction. Their circulation is quantised, protecting them against decay. This section discusses the basic concepts and phenomena of these vortex lines.

5.1 Circulation quantisation

The relationship (1.14) between the superfluid velocity $\mathbf{v}_s$ and the wavefunction phase θ has an important physical significance for rotating fluids. Consider any closed loop C in the fluid. The circulation around this loop is defined as the line integral of the velocity around the loop:

$$K = \oint_C \mathbf{v}_s \cdot d\mathbf{r} \tag{1.25}$$

In terms of the phase θ,

$$K = \frac{\hbar}{m} \oint_C \boldsymbol{\nabla}\theta \cdot d\mathbf{r} = \frac{\hbar}{m}\Delta\theta \tag{1.26}$$

where $\Delta\theta$ is the difference in phase between the beginning and the end of the path. Since the many-body wavefunction Ψ and the phase $\theta(\mathbf{r})$ must be single-valued, $\Delta\theta$ around the loop must be zero or an integer multiple of 2π. Therefore, the circulation is quantised in units of h/m:

$$K = \oint_C \mathbf{v}_s \cdot d\mathbf{r} = \frac{h}{m} n_K \qquad \text{for an integer } n_K \tag{1.27}$$

Since the phase of the condensate wavefunction is changing as we go around the loop, there must be fluid flow with velocity $\mathbf{v}_s \propto \boldsymbol{\nabla}\theta$ around the loop. Moreover, the phase changes by $2\pi n_K$ around the loop. This means that the condensate wavefunction carries angular momentum $\hbar n_K$ around an axis which threads the loop. There are N bosons condensed in this wavefunction, and so we have a total angular momentum of

$$L_z = N n_K \hbar \tag{1.28}$$

In other words, circulation quantisation is a **macroscopic manifestation of the quantisation of angular momentum**.

The fluid with finite circulation contains **vortex lines**. Suppose we have a superflow which is cylindrically symmetrical around the z-axis, *i.e.* v_ϕ is only a function of r. Then, the circulation around a circular path of radius r centred on the z-axis is $K = v_\phi(r) \times 2\pi r$. If we have a non-zero circulation $K = n_K h/m$ around the z-axis, then the tangential velocity is $v_\phi(r) = n_K \hbar/mr$. This vortex line apparently has infinite fluid velocity as we approach the z-axis ($r \to 0$). In other words, we have a *line singularity* in the superfluid.

The singularity with infinite velocity is of course unphysical. The high tangential velocity gives a centrifugal force which flings the particles outwards from the z-axis. In fact, the superfluid density vanishes along the z-axis (see Fig. 1.8) so that the velocity is not defined along the axis of this vortex line. This is how we manage to escape the conclusion that $\boldsymbol{\nabla} \times \mathbf{v}_s = 0$ must be zero everywhere in a flow field: there is a line where neither θ nor $\mathbf{v}_s$ is defined at the core of the vortex. Physically, this creates a non-superfluid **normal core** around the z-axis where the condensate density is reduced from the bulk value. The core size is of the order of the healing length. The quantisation of angular momentrum means that these are **topologically stable** configurations of the system. We cannot undo the 2π phase change around a vortex by reducing it gradually to zero. Otherwise, we would have a wavefunction that is not single-valued in space. The only way in which a vortex can be annihilated is to meet an antivortex, *i.e.* a vortex with a circulation of the opposite sign.

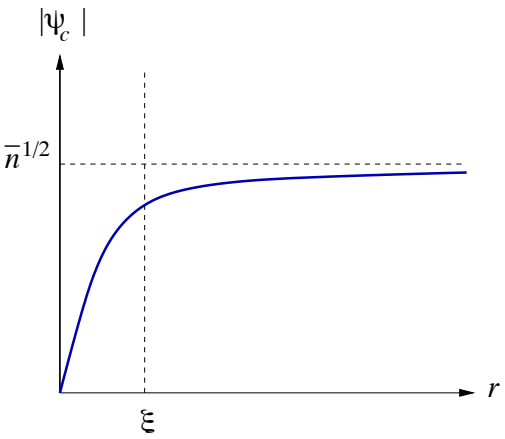


Figure 1.8: Schematic diagram of condensate wavefunction amplitude for a vortex showing normal core within a healing length ξ from the axis. Far away from the core ($r \gg \xi$), the wavefunction approaches its asymptotic value of $\sqrt{\bar{n}}$ as a power law.

5.2 Vortex line energy

Consider a straight vortex line with its axis along the axis of symmetry of a cylinder of radius R and length L. Suppose it has a circulation of $K = n_K h/m$. Let us use cylindrical polars: r for the distance from the axis of the vortex and ϕ for the azimuthal angle around the vortex.

The condensate phase is $\theta = n_K \phi$ (plus an arbitrary constant), and so the tangential velocity is $v_\phi = n_K \hbar/mr$. As we see from Fig. 1.8, we can take the density to be approximately constant everywhere except in the normal core. The vortex does not alter the repulsive energy of the system. The kinetic energy dominates the total energy:

$$E_V = \int \frac{m\mathbf{v}_s^2}{2}\, n(\mathbf{r}) d^3\mathbf{r} = \frac{\bar{n} m L}{2} \int_\xi^R v_\phi^2\, 2\pi r\, dr = n_K^2 L\, \frac{\pi\hbar^2 \bar{n}}{m} \ln\left(\frac{R}{\xi}\right) \tag{1.29}$$

where we have approximated the boson number density n by its asymptotic value $\bar{n}$ away from the core. We should add a contribution from the core. Its exact value of the correction is sensitive to the core structure.

Note also that the energy per unit length of the vortex (the line tension E_V/L) is proportional to the square of the circulation n_K^2: this means that it is energetically favorable to have two singly-quantised vortices ($1^2 + 1^2 = 2$) rather than one doubly-quantised vortex ($2^2 = 4$).

Moreover, this line tension diverges logarithmically when the radial size of the system is taken to infinity ($R/\xi \to \infty$). This means that, even for a thin film where vortex lines can be short, the energy cost is prohibitively high and single vortex lines are not thermally excited. Examining (1.29), we can see that this logarithmic divergence can be traced back to fluid flow *far away* from the vortex axis. Although the fluid speed decays as $1/r$, this decay is not fast enough and distant fluid motion dominates the energy cost.

On the other hand, if we have a vortex-antivortex pair with n_K of opposite signs, the $1/r$ tail of their flow fields cancel and the divergent energy cost is cured. For a vortex-antivortex separation of d, the energy cost per unit length is:

$$\frac{E_{V\bar{V}}(d)}{L} = \frac{2\pi\hbar^2\bar{n}}{m} \ln\left(\frac{d}{\xi}\right) \tag{1.30}$$

The logarithmic dependence remains, but it is for the separation of the vortices rather

than the system size R. Vortex-antivortex pairs play a specially important role in the behaviour of superfluid thin films.

5.3 Rotating superfluids

Vortices are the only objects which support angular momentum in the bulk of the superfluid. Suppose we have a rotating cylinder of liquid helium above the superfluid transition temperature T_c. It will be rotating rigidly with the wall of the container. If we cool it below T_c, the flow pattern alters drastically (keeping angular momentum constant) so that the angular momentum is bundled into discrete vortex lines in the fluid. The number of vortex lines will increase proportionally with the angular momentum of the fluid. In this section, I will discuss the energetics of this rotating superfluid.

To understand this, we perform a rotational version of the analysis that led us to the Landau criterion for a critical velocity for linear motion. When we start rotating the walls of the cylinder, the relative motion of the wall and superfluid means that any small perturbations on the wall can impart angular momentum to the fluid. It is therefore sensible to consider the frame that rotates with the cylinder wall. A rigid rotation only affects the kinetic part of the energy. The velocity of each particle is measured in the rotating frame as $\mathbf{v}' = \mathbf{v} - \boldsymbol{\omega} \times \mathbf{r}$. Using this, it can be shown that energy changes $\Delta E'$ in the rotating frame is related to energy changes ΔE in the fluid frame by

$$\Delta E' = \Delta E - \boldsymbol{\omega} \cdot \Delta \mathbf{L} \tag{1.31}$$

where $\Delta \mathbf{L}$ is the change in total angular momentum from the ground state (which usually has zero angular momentum). The first vortex line will enter the system if $\Delta E'$ is negative. For a straight vortex line, this occurs at a critical angular velocity ω_{c1} of

$$\omega_{c1} = \frac{\hbar}{mR^2} \ln\left(\frac{R}{\xi}\right) \tag{1.32}$$

for the first vortex line to enter the fluid. [**Exercise**: show this using the expressions for energy (1.29) and angular momentum (1.28) of the vortex line.]

If ω is below this threshold value, the fluid does not rotate with the cylinder at all! This is a direct consequence of superfluidity. However, this regime of no rotation is difficult to observe: the threshold ω_{c1} for vortex penetration is very low for a macroscopically large system ($R \gg \xi$).

Above this threshold, vortex lines enter the cylinder. For a given angular frequency ω, the vortices will enter until there is no relative motion between the fluid and the container wall. These vortex lines form a triangular array of lines to minimise their interactions (see Fig. 1.9).

As we increase the angular velocity, the array becomes denser and denser as more vortex lines enter the system. Superfluidity is destroyed when the lines are so dense that their normal cores overlap. The radius of the vortex core is approximately the healing length ξ. When the vortex cores overlap, there are $1/\xi^2$ vortices per unit cross-sectional area of the triangular lattice, giving a total circulation of $K = (h/m)(\pi R^2/\xi^2)$ around the wall of the cylinder. This corresponds to a velocity of $v_\phi = K/2\pi R = (h/2m)(R/\xi^2)$ at the wall. If there is no relative motion between the wall and the fluid, it is rotating at an angular

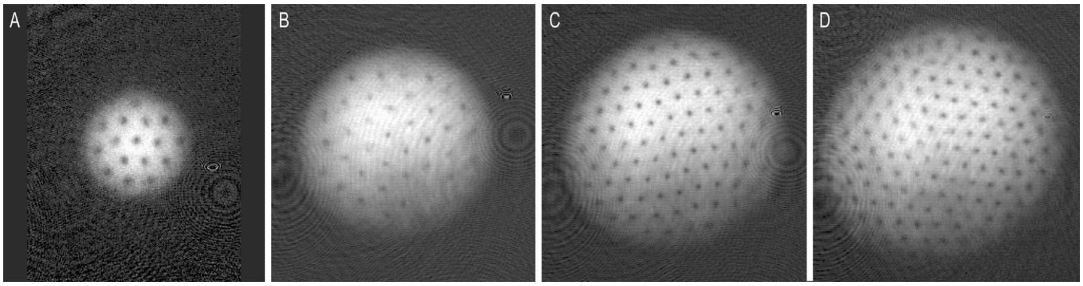


Figure 1.9: Triangular vortex lattice in a rotating Bose condensate of ultracold Na atoms. These examples contain approximately (A)16, (B)32, (C)80, (D)130 vortices. [J.R. Abo-Shaeer, C. Raman, J.M. Vogels and W. Ketterle, *Science* **292** 476 (2001)]

frequency of $v_\phi/R = (h/2m\xi^2)$. So, we estimate that

$$\omega_{c2} \sim h/m\xi^2 \tag{1.33}$$

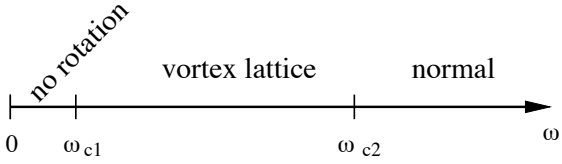


Figure 1.10: Phases of a rotating Bose fluid as a function of angular frequency ω.

6 Dynamics

6.1 Gross-Pitaevskii equation

To study the dynamics of the condensate, we need to introduce time dependence into the nonlinear Schrödinger equation. This approach is suitable near zero temperature. At higher temperatures, there will be significant contribution from the normal viscous fraction of the fluid. The dynamics of the dynamics is described by the Gross-Pitaevskii (GP) equation:

$$i\hbar\frac{\partial}{\partial t}\psi_{\rm c} = -\frac{\hbar^2}{2m}\nabla^2\psi_{\rm c} + u(|\psi_{\rm c}|^2 - \bar{n})\psi_{\rm c} \tag{1.34}$$

This is an obvious choice. First of all, it works in the non-interacting limit when each particle occupies the same time-evolving single-particle state. It can also be justified in terms of a variational principle for the action:

$$S = \int dt \int d^3\mathbf{r}_1 \ldots d^3\mathbf{r}_N \left[i\hbar\Psi^*\frac{\partial\Psi}{\partial t} - \Psi^* H \Psi \right]$$

In other words, the above equation is obtained if we restrict the minimisation to totally condensed states. [**Exercise**: try it.]

Before proceeding to make use of the GP equation, we should check it obeys fundamental conservation laws. All fluid flow conserves matter locally: if particles have entered or left a

region, then they must have come from or gone to a neighbouring region. This conservation law is contained in the **continuity equation** which relates the rate of change of density n with the divergence of current density $\mathbf{J}$:

$$\frac{\partial n}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{J} = 0 \tag{1.35}$$

In terms of the condensate wavefunction $\psi_{\rm c}$, we have $n = |\psi_{\rm c}|^2$ and so, using the GP equation (1.34), we see that

$$\frac{\partial n}{\partial t} = \psi_{\rm c}^* \frac{\partial \psi_{\rm c}}{\partial t} + \psi_{\rm c} \frac{\partial \psi_{\rm c}^*}{\partial t} = \frac{i\hbar}{2m}(\psi_{\rm c}^* \nabla^2 \psi_{\rm c} - \psi_{\rm c} \nabla^2 \psi_{\rm c}^*)$$

This is indeed equal to the current density (1.13):

$$\boldsymbol{\nabla} \cdot \mathbf{J} = \frac{-i\hbar}{2m} \boldsymbol{\nabla} \cdot (\psi_{\rm c}^* \boldsymbol{\nabla} \psi_{\rm c} - \psi_{\rm c} \boldsymbol{\nabla} \psi_{\rm c}^*) = \frac{-i\hbar}{2m}(\psi_{\rm c}^* \nabla^2 \psi_{\rm c} - \psi_{\rm c} \nabla^2 \psi_{\rm c}^*)$$

6.2 Hydrodynamics

To see what the dynamical equation (1.34) means for the density and current in the superfluid, let us multiply (1.34) by $\psi_{\rm c}^*$ to obtain:

$$i\hbar \psi_{\rm c}^* \partial_t \psi_{\rm c} = -(\hbar^2/2m)\psi_{\rm c}^* \nabla^2 \psi_{\rm c} + u\psi_{\rm c}^*(|\psi_{\rm c}|^2 - \bar{n})\psi_{\rm c} \tag{1.36}$$

Let us now substitute for the density and phase variables: $\psi_{\rm c} = \sqrt{n} e^{i\theta}$. Since n and θ are real variables, this gives an equation with explicitly real and imaginary parts. The imaginary part of the equation gives the continuity equation again. [**Exercise**: check this.] After some simple manipulation, the real part of the equation gives a dynamical equation for the phase:

$$-\hbar \frac{\partial \theta}{\partial t} = \frac{\hbar^2}{2m} |\nabla \theta|^2 + u(n - \bar{n}) - \frac{\hbar^2}{2m\sqrt{n}} \nabla^2 \sqrt{n} \tag{1.37}$$

To understand the physical origin of this equation, let us write this equation in terms of the superfluid velocity $\mathbf{v}_s = \hbar \boldsymbol{\nabla} \theta / m$, rather than the condensate phase. We can write the right-hand side of the equation for $\partial\theta/\partial t$ in terms of density and velocity:

$$-\hbar \frac{\partial \theta}{\partial t} = \frac{1}{2} m \mathbf{v}_s^2 + u(n - \bar{n}) - \frac{\hbar^2}{2m\sqrt{n}} \nabla^2 \sqrt{n} \tag{1.38}$$

To put the left-hand side in terms of the velocity, we take the gradient of this equation. We need the identity: $\boldsymbol{\nabla}(\mathbf{v}^2)/2 = (\mathbf{v} \cdot \boldsymbol{\nabla})\mathbf{v} + \mathbf{v} \times (\boldsymbol{\nabla} \times \mathbf{v})$. Our current derivation uses $\mathbf{v}_s \propto \boldsymbol{\nabla}\theta$ which has no vorticity ($\boldsymbol{\nabla} \times \mathbf{v}_s = 0$), so the second term in this identity vanishes[7]. After some rearrangement, we obtain an equation for the acceleration of the fluid:

$$m \frac{d\mathbf{v}_s}{dt} \equiv m \left(\frac{\partial}{\partial t} + \mathbf{v}_s \cdot \boldsymbol{\nabla} \right) \mathbf{v}_s = -\boldsymbol{\nabla} \mu \tag{1.39}$$

$$\mu(\mathbf{r}) = un - (\hbar^2/2m\sqrt{n})\, \nabla^2 \sqrt{n} \tag{1.40}$$

[7] We can add a singular part to the phase function to include vorticity. The hydrodynamic equation (1.39) holds in general.

where $d\mathbf{v}_s/dt$ denotes the acceleration of a fluid particle as it moves through the fluid. This derivative is not the time derivative of the velocity at a fixed point in space. Instead, it follows the particle as it moves: $d/dt = \partial_t + (d\mathbf{r}/dt) \cdot \partial_{\mathbf{r}} = \partial_t + \mathbf{v} \cdot \boldsymbol{\nabla}$.

Equation (1.39) is nothing but Newton's second law for the fluid particles moving in a spatially inhomogeneous local chemical potential $\mu(\mathbf{r})$. The statement of energy conservation along a stream line is:

$$\frac{1}{2}m\mathbf{v}_s^2 + \mu(\mathbf{r}) = \text{constant} \tag{1.41}$$

Replacing μ by p/n gives us the conventional form of **Bernouilli's equation** and so we can identify $n\mu$ as a pressure.

6.3 Sound waves

We will now consider now small perturbations around the ground state and see how the system evolves. The ground state of the system has $\psi_{\rm c} = \sqrt{\bar{n}}$. Let us assume that the density and phase fluctuations are small: $|\delta n| \ll \bar{n}$ and $|\theta(\mathbf{r}, t)| \ll 2\pi$. So, the condensate wavefunction is

$$\psi_{\rm c}(\mathbf{r}, t) = \sqrt{\bar{n}}\,(1 + \delta n/\bar{n})^{1/2} e^{i\theta}$$

Following our approach for the static NLSE, we can express the problem in natural lengths and time. The natural length is the healing length $\xi = (\hbar^2/mu\bar{n})^{1/2}$ as discussed before (1.24). The natural energy scale is the Hartree energy $u\bar{n}$ so the natural time scale is $\hbar/u\bar{n}$. With ξ as unit of length and $\hbar/u\bar{n}$ as unit of time, we can define dimensionless space and time coordinates $\mathbf{a} = \mathbf{r}/\xi$ and $\tau = u\bar{n}t/\hbar$. Let the fractional fluctuations in the density be $\nu = (n - \bar{n})/\bar{n}$, then we can write the hydrodynamic equations (1.37) as

$$\begin{aligned} -\frac{\partial\theta}{\partial\tau} &= \frac{1}{2}|\nabla_{\mathbf{a}}\theta|^2 + \nu - \frac{1}{2\sqrt{1+\nu}}\nabla_{\mathbf{a}}^2\sqrt{1+\nu} \\ \frac{\partial\nu}{\partial\tau} &= -\nabla_{\mathbf{a}} \cdot [(1+\nu)\nabla_{\mathbf{a}}\theta] \end{aligned} \tag{1.42}$$

For $|\nu|$ and θ small, we can obtain an equation linear in ν and θ by dropping higher order terms (*e.g.*, ν^2, $|\boldsymbol{\nabla}\theta|^2$):

$$-\frac{\partial\theta}{\partial\tau} = \left(1 - \frac{1}{4}\nabla_{\mathbf{a}}^2\right)\nu \qquad \text{and} \qquad \frac{\partial\nu}{\partial\tau} = -\nabla_{\mathbf{a}}^2\theta \tag{1.43}$$

Taking the time derivative of the first equation and substituting for $\partial\nu/\partial\tau$ using the second (continuity) equation,

$$\frac{\partial^2\theta}{\partial\tau^2} = \left(1 - \frac{1}{4}\nabla_{\mathbf{a}}^2\right)\nabla_{\mathbf{a}}^2\theta$$

Restoring the original units, we have the equation of motion for the condensate phase $\theta(\mathbf{r}, t)$:

$$\frac{\partial^2\theta}{\partial t^2} = \frac{u\bar{n}}{m}\left(1 - \frac{\xi^2}{4}\nabla^2\right)\nabla^2\theta \tag{1.44}$$

Therefore, the phase of the condensate propagates as a wave: $\theta \sim \theta_{\mathbf{k}} \exp[i(\mathbf{k} \cdot \mathbf{r} - \omega_{\mathbf{k}} t)]$ with the frequency spectrum:

$$\omega_{\mathbf{k}}^2 = \frac{u\bar{n}}{m}k^2\,(1 + k^2\xi^2) = \frac{k^2}{2m}\left(2u\bar{n} + \frac{\hbar^2k^2}{2m}\right) \tag{1.45}$$

Notice that the frequency goes to zero in the limit of very long wavelengths ($k \to 0$). This is expected (and is a good check) because $k = 0$ corresponds to a globally uniform phase change to the wavefunction; this has no effect on the energy of the system whatsoever.

To get a more physical picture of these propagating waves, we can see how the density and velocity behaves. The density fluctuations can be obtained from the phase fluctuations through the continuity equation in its linearised form (1.43), while the velocity can be obtained from the gradient of the phase (1.14). We see that

$$\begin{aligned} \theta &= \theta_{\mathbf{k}} \sin(\mathbf{k}\cdot\mathbf{r} - \omega_{\mathbf{k}} t) \\ \Rightarrow \mathbf{v}_s &= \frac{\hbar\mathbf{k}}{m}\,\theta_{\mathbf{k}}\,\cos(\mathbf{k}\cdot\mathbf{r} - \omega_{\mathbf{k}} t)\ , \quad \delta n = \bar{n}\nu = \frac{\bar{n}k^2}{\omega_{\mathbf{k}}}\,\theta_{\mathbf{k}}\,\cos(\mathbf{k}\cdot\mathbf{r} - \omega_{\mathbf{k}} t) \end{aligned}$$

We see that the fluid flow is parallel to the direction of wave propagation, and the density and current are in phase: this wave is a *longitudinal* density wave, *i.e.* a **sound wave**. From the dispersion relation (1.45), we see that, at wavelengths long compared to the healing length ξ, we have a non-dispersive wave which has a sound velocity $c_s = (u\bar{n}/m)^{1/2}$. Put another way, if we drop the double Laplacian term in the differential equation (1.44), the equation becomes the conventional wave equation.

Note that the sound velocity is not a quantum mechanical quantity — no $\hbar$. Indeed, any classical fluid with a finite compressibility would have a longitudinal sound wave. Quantum physics enters the picture because these collective modes of the fluid become quantised to excitations with energy $\epsilon_{\mathbf{k}} = \hbar\omega_{\mathbf{k}}$.

7 Superfluidity, spontaneous symmetry breaking and Goldstone modes

As we discussed in section 2.1, the Landau criterion tells us that the excitation spectrum of the system has important consequences for the existence of superfluidity in the system. Recall that the quadratic dispersion relation of the excitations of the ideal Bose gas was fatal to its superfluidity: Landau's criterion (1.9) says that the condensate is destroyed at any flow velocity. From (1.45), we see that the presence of interactions changes the quadratic dispersion to a linear one at wavelengths larger than the healing length. An excitation in the Bose fluid with momentum $\Delta\mathbf{P} = \hbar\mathbf{k}$ has energy $\Delta E = \hbar\omega_{\mathbf{k}} \propto |\mathbf{k}|$. Upon a Galilean transformation $\epsilon_{\mathbf{k}} \to \epsilon_{\mathbf{k}} - \hbar\mathbf{v}\cdot\mathbf{k}$, the excitation energies remain positive as long as the relative velocity of the new frame $|\mathbf{v}|$ is less than the sound velocity c_s (Fig. 1.5). Therefore, this suggests that **the interacting Bose liquid is a superfluid**.

This conclusion is correct provided that there are *no other low-energy excitations that fail Landau's criterion*. This can be proved in two dimensions and above using a sum rule but the proof is beyond the scope of these lectures[8]. In fact, this is a very general consequence of the fact that the ground state breaks a global continuous symmetry of the Hamiltonian. In the case of superfluids, this continuous symmetry is the fact that the energy of the system does *not* depend the overall value of the condensate phase. In other words, the continuous transformation:

$$\theta(\mathbf{r}) \to \theta(\mathbf{r}) + \phi\ , \qquad \psi_{\rm c} \to \psi_{\rm c} e^{i\phi} \tag{1.46}$$

[8]This does not apply in one dimension. There are in fact gapless excitations near the wavevector $q = 2\pi n$.

does not affect the energy (1.19) at all for any constant shift ϕ. In spite of this invariance, the ground-state condensate wavefunction takes on a definite phase (conventionally taken to be zero). This global phase does not fluctuate in time over all possible values between 0 and 2π. (Otherwise, the complex order parameter ψ_c would have averaged to zero and we could not describe the system in terms of a simple condensate wavefunction.) In fact, the phase does have fluctuations even at zero temperature. However, these quantum fluctuations are small provided that the interactions are weak or the density is low.

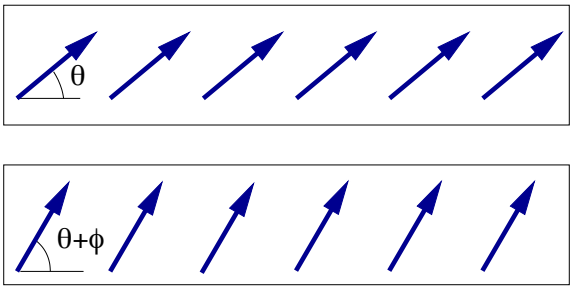


Figure 1.11: Two ground states with spatially uniform phases, θ and $\theta+\phi$, for the complex order parameter ψ_c. The direction of the arrow represents the direction of ψ_c in the complex plane.

This phenomenon of the system choosing a definite phase is called **spontaneous symmetry breaking**. This choice is 'spontaneous' because there is no energetic reason to choose one value of the phase over another (Fig. 1.11). A more concrete example is the isotropic ferromagnet: although the magnetic moment of a ferromagnet can point anywhere, the ground state of the magnet has a magnetisation in a definite direction (even in the absence of a magnetic field aligning the spins *via* the Zeeman effect). In the case of superfluids, the condensate phase plays an analogous role to the magnetisation direction of a ferromagnet.

The condensate wavefunction ψ_c of the superfluid and the magnetisation $\mathbf{M}$ of the ferromagnet are called the **order parameters** associated with symmetry breaking. In both systems, the Hamiltonian of the system is invariant under a global transformation of the order parameter. For magnets, the transformation is a global rigid rotation of all the spins. Mathematically, this is a SU(2) rotation in spin space. For superfluids, it is a global rotation ψ_c in the complex plane (see Fig. 1.11), belonging to the U(1) group.

In both examples of the superfluid and the isotropic ferromagnet, the symmetry of the Hamiltonian involves **continuous** transformations of the order parameter. This should be contrasted with systems with *discrete* symmetries, such as the Ising ferromagnet where the magnetisation can only point in the $+z$ or $-z$ directions. In particular, **Goldstone's theorem** states that, if a *continuous* global symmetry is spontaneously broken in a system, then the system will have **gapless excitations** at long wavelengths, meaning that $\omega_{\mathbf{k}} \to 0$ as $k \to 0$. These excitations consist of long-wavelength twists in the phase for a U(1) order parameter of a superfluid, or smoothly varying SU(2) rotations of the magnetisation). These are called **Goldstone modes**.

We have already seen the Goldstone modes of a superfluid explicitly from the time-dependent Schrödinger equation. They are just the sound waves. In terms of the condensate phase, the energy of such an excitation vanishes in the long-wavelength limit essentially because the effect on the order parameter becomes closer and closer to a rigid global phase shift and, from the symmetry of the Hamiltonian, we know that this should not cost any energy in the limit of taking the wavelength to infinity.

However, this is not the whole story — Goldstone's theorem also guarantees that these

excitations are **long-lived**, *i.e.* they do not decay in the limit of large wavelengths[9]. This implies that there are no other excitations at the same energy or below into which the sound wave can decay. This is the key ingredient that allows the interacting Bose fluid to satisfy Landau's criterion for *all* low-energy excitations.

[9] J. Goldstone, A. Salam and S. Weinberg, *Phys. Rev.* **127** 965 (1962).

Chapter 2

Charged superfluids

- Phenomenology of charged superfluids
  - electron pairing → Cooper pairs → charged superfluids
  - Ginzburg-Landau theory
  - Meissner effect: magnetic flux expulsion
  - Flux quantisation: magnetic flux quantised in units of the flux quantum $h/2e$
  - two types of superconductors
- Josephson effect: dc effect
  - SIS junction for pair tunnelling
  - phase difference between two superconductors → pair tunnelling current: $I = I_c \sin(\Delta\theta)$
- Macroscopic Aharonov-Bohm effect
  - superconducting ring with one junction
  - SQUID ring: Young's slits for Cooper pairs

1 Overview

Superconductors are the charged cousins of neutral superfluids. They have zero electrical resistance below a critical temperature T_c. They also expel magnetic fields (Meissner effect). They do so by generating circulating persistent currents which induce a magnetic field to cancel out the externally applied field.

These materials have important technological applications. Zero electrical resistance allows efficient energy storage and power transmission. We can also use superconducting wires to generate large magnetic fields. Such superconducting magnets can be used in large-scale applications, such as particle accelerators (Large Hadron Collider), or in equipment at a smaller scale, such as magnetic resonance imaging (MRI). Moreover, the dramatic response of superconductors to magnetic fields means that we can use it as a sensitive probe of these fields.

The first superconductor was discovered by Kamerlingh Onnes in 1911, soon after the liquefaction of helium became possible. He noticed that the resistance of liquid mercury vanished at 4K. In fact, most elements become superconducting at low temperatures, although some have to be encouraged to do so by high pressure (Fig. 2.1).

Superconductivity is not restricted to elements. Many compounds and alloys also superconduct. Nb_3Sn ($T_c = 18$K) is a well-studied material in the context of applications. Among the superconductors that can be explained by conventional theory (using an

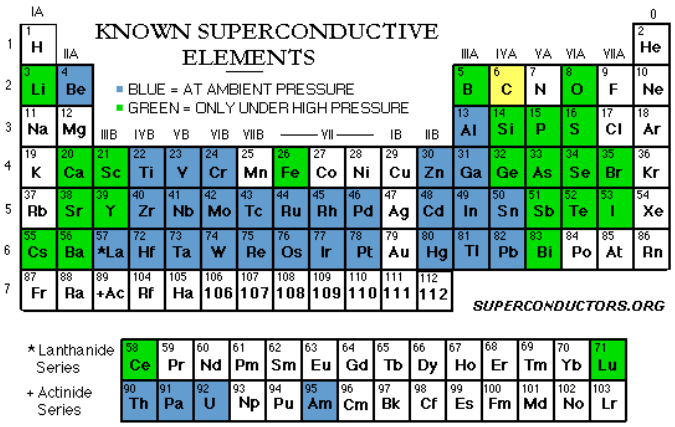


Figure 2.1: Periodic table of superconducting elements. Carbon: superconducts as single-walled carbon nanotubes. [http://www.superconductors.org]

electron-phonon mechanism), MgB_2 probably has the highest T_c (39K). Strangely, although this compound is easy to synthesise and commonly found in chemistry labs, its superconductivity was only discovered serendipitously[1] by a Japanese group looking for unconventional magnetism. New superconductors continue to be discovered. The most recent surge of interest came from the 2008 discovery of iron-based pnictides, such as $LaO_{1-x}F_xFeAs$ ($T_c = 26$K)[2] which consists of layers of FeAs and LaO with F substituted to around 10%. Other classes of compounds that exhibit superconductivity include the intercalated graphite compounds[3] (CaC_6, YbC_6 with T_c of 11.5K and 6.5K respectively) and alkali-doped fullerenes[4] (A_3C_{60} with A=K,Rb,Cs where $T_c = 18$, 30 and 40K).

The highest known superconducting critical temperatures come from a class of compounds known as the cuprate superconductors. The first cuprate superconductor, $La_{2-x}Sr_xCuO_4$ with a T_c of 40K, was discovered in 1986 by Bednorz and Müller[5]. It became evident that a whole class of such copper-oxide compounds would superconduct when $YBa_2Cu_3O_7$ was found by Paul Chu's group[6] to superconduct at 90K. Although room temperature would be most desirable, an important benchmark is a critical temperature above the liquid nitrogen boiling point (77K) because this allows us to use commonly available liquid nitrogen as a coolant, instead of very expensive liquid helium. The record for the superconducting transition temperature is 138K for $HgBa_2Ca_2Cu_3O_8$, but its mercury content makes its preparation process highly toxic. There has also been a recent report of a 150K superconductor ($InSnBa_4Tm_3Cu_5O_x$).

The absence of resistance seems to indicate that some form of superfluidity is responsible for superconductivity. However, the mobile charge carriers in metals are electrons: they are fermions and so the Pauli exclusion principle prevents them from undergoing Bose condensation. Nature circumvents this problem by pairing electrons into charge-$2e$ bosons known as **Cooper pairs**. The magnetic response of superconductors can be understood

[1] J. Nagamatsu *et al.*, *Nature*, **410**, 63 (2001).
[2] Y. Kamihara *et al.*, *J. Am. Chem. Soc.* **130**, 3296 (2008).
[3] T.E. Weller *et al.*, *Nature Physics*, **1**, 39 (2005).
[4] A.F. Hebard *et al*, *Nature*, **350**, 600 (1991).
[5] J.G. Bednorz and K.A. Müller, *Z. Phys. B Condensed Matter*, **64**, 189 (1986).
[6] M.K. Wu *et al.*, *Phys. Rev. Lett.*, **58**, 908 (1987).

to a large degree by modelling the system as a superfluid of these Cooper pairs. This is the *Ginzburg-Landau theory* that we discuss in this chapter. This theory provides a phenomenological framework for the macroscopic behaviour of superconductors. In fact, this phenomenology was well established before a microscopic theory for Cooper pairs was put forward by Bardeen, Cooper and Schrieffer (BCS). Indeed, we still do not have a microscopic theory for the cuprate superconductors.

2 Phenomenological theory

In this section, I will describe a simple phenomenological description of Cooper pairs as charge-$2e$ bosons by adapting the theory for neutral bosons. Let us start with the case of no magnetic fields ($\mathbf{B} = 0$). We need an analogue of the variational treatment that leads to the non-linear Schrödinger equation (NLSE) (1.22). We can write down a variational energy for the Cooper pair condensate of the form:

$$F = \int d^3\mathbf{r} \left[\frac{\hbar^2}{2m^*} |\nabla \psi_{\rm c}|^2 - \mu |\psi_{\rm c}|^2 + \frac{u}{2} |\psi_{\rm c}|^4 \right] \qquad \text{for } \mathbf{B} = 0$$

For the moment, we can simply treat m^*, μ and u as phenomenological parameters to be fitted to experiments. It is natural to expect that, at zero temperature, $m^* \simeq 2m_e$ is the effective mass for the Cooper pair.

However, this analogy with neutral superfluids is really only appropriate if the electrons are so tightly bound in the Cooper pairs that they form a distinct bosonic entity. Unfortunately, most superconductors are not in this strong-binding limit. Nevertheless, this phenomenlogical theory manages to produce the basic phenomena of superconductivity.

In fact, this form of the energy F is well known in the **Ginzburg-Landau theory** of thermal phase transitions as the *free energy* of the system at a finite temperature. In such a theory, the quantity $\psi_{\rm c}$ is treated as an **order parameter** that measures the degree of superfluidity or condensation. An order parameter represents the presence of order in the system, such as spontaneous magnetisation, Bose condensation, superconductivity, crystalline order, *etc.* At this level of treatment, the only aspect of the theory that distinguishes between different physical phenomena is the nature of the order parameter. For superfluids and superconductors, it is a complex number [$U(1)$ symmetry]. For an isotropic ferromagnet, it should be a three-dimensional vector [$O(3)$ symmetry] which gives the magnetisation of the system. Landau's rationale behind using the above form for the free energy is that, close to the critical temperature, the free energy should be an analytic function of $\psi_{\rm c}$ and can be approximated by a Taylor expansion around $\psi_{\rm c} = 0$. He parametrised this as a simple linear function

$$\mu(T) \propto (T_c - T) \tag{2.1}$$

which changes sign when the temperature goes below T_c. Minimising this Landau form of the free energy for a spatially constant $\psi_{\rm c}$, we see that the tuning of $\mu(T)$ through zero marks the onset of order in the system:

$$|\psi_{\rm c}| = \begin{cases} (\mu/u)^{1/2} \propto (T_c - T)^{1/2} & \text{for } T < T_c \\ 0 & \text{for } T > T_c \end{cases} \tag{2.2}$$

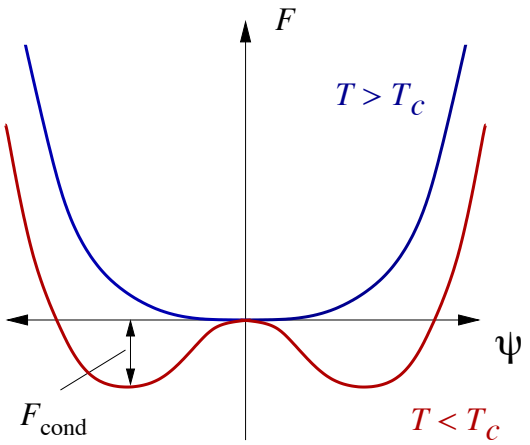


Figure 2.2: Landau free energy. A minimum develops at a finite value of ψ_c below a critical temperature T_c. The condensation energy $F_{\rm cond}$ is the depth of this minimum.

Below the critical temperature, the superconducting state has a lower free energy than the normal state ($|\psi_c| = 0$). The difference is the **condensation energy** of the system. In these parameters, the condensation energy per unit volume is

$$f_{\rm cond} = f_{\rm normal} - f_s = u|\psi_c|^4/2 = \mu^2/2u \tag{2.3}$$

2.1 Ginzburg-Landau theory

Now, we can generalise the theory to include an external magnetic field: one simply adds a coupling of the vector potential to the canonical momentum of the charge-$2e$ Cooper pair[7]. So, the kinetic energy for the supercurrents becomes $\hbar^2|\boldsymbol{\nabla}\psi_c|^2/2m^* \to |(-i\hbar\boldsymbol{\nabla} - 2e\mathbf{A})\psi_c|^2/2m^*$. This leads us to the total variational energy for superconductors in a static magnetic field:

$$F_{\rm GL}[\psi_c, \mathbf{A}] = \int d^3\mathbf{r} \left[\frac{1}{2m^*}|(-i\hbar\boldsymbol{\nabla} - 2e\mathbf{A})\psi_c|^2 - \mu|\psi_c|^2 + \frac{u}{2}|\psi_c|^4 + \frac{1}{2\mu_0}\mathbf{B}^2\right] \tag{2.4}$$

with $\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}$. Note that, in addition to the energy of the superconducting Cooper pairs, we also have to include the energy in the magnetic fields in order to obtain the correct physical behaviour of the whole electrodynamical system. This means that $F_{\rm GL}$ is a functional of *both* the condensate wavefunction ψ_c and the vector potential $\mathbf{A}$, and we should minimise this energy with respect to both fields. Physically, this is because an externally applied field will induce supercurrents which produce internally induced magnetic fields, and these fields cost energy to generate.

Before proceeding to discuss the magnetic response of superconductors, let us recall that neutral superfluids have a natural length scale, the healing length (1.24). At zero magnetic field, the Ginzburg-Landau theory gives us an analogous characteristic length:

$$\xi = (\hbar^2/2m^*u|\psi_c|^2)^{1/2} \tag{2.5}$$

This is called the **coherence length** in superconductivity[8]. As already mentioned before, Cooper pairs can be quite extended objects, and so it is not appropriate to think of this

[7] It seems to be the convention that we hide the negative sign of the electron charge in e.

[8] I have introduced a factor of two in the denominator to be consistent with convention in superconductivity literature.

directly as the healing length of a superfluid of charged bosons. In fact, for a clean superconductor, the coherence length is approximately the size of the Cooper pair. Note that, as T approaches T_c, the order parameter vanishes (2.2), and so the coherence length diverges:

$$\xi(T) \sim (T_c - T)^{-1/2} \qquad \text{as } T \to T_c \tag{2.6}$$

We will see below that a characteristic magnetic length scale arises from this theory too. The emergence of this new magnetic length makes superconductors quite different from neutral superfluids in many respects.

3 Magnetic Response

3.1 London equation and Meissner effect

We do not need the full Ginzburg-Landau theory to understand the basic magnetic response of a superconductor. In particular, if we work at temperatures not too close to T_c and at relatively weak magnetic fields, we can assume that the magnitude of the order parameter $|\psi_c|$ is constant (at length scales larger than the coherence length ξ). Then, we have only one variable left: the phase θ. We can write $\psi_c(\mathbf{r}) = \sqrt{n_s/2}e^{i\theta(\mathbf{r})}$ where

$$n_s = 2|\psi_c|^2 \tag{2.7}$$

is interpreted[9] as a density of superconducting electrons and $|\psi_c|^2$ is the density of pairs. Also, the current density of the charged condensate is given by:

$$\mathbf{J} = (2e)|\psi_c|^2\mathbf{v} = n_s e\mathbf{v} = \frac{n_s e}{m^*}(\hbar\boldsymbol{\nabla}\theta - 2e\mathbf{A}) \tag{2.8}$$

In the approximation of constant $|\psi_c|$ over length scales larger than the coherence length, the variational energy (2.4) simplifies to:

$$F[\theta, \mathbf{A}] = \int d^3\mathbf{r}\left[\frac{n_s}{4m^*}(\hbar\boldsymbol{\nabla}\theta - 2e\mathbf{A})^2 + \frac{1}{2\mu_0}\mathbf{B}^2\right] \tag{2.9}$$

We can write the variational energy totally in terms of electromagnetic quantities. Using (2.8), the kinetic energy density is simply $n_s m_e \mathbf{v}_s^2/2 = m^*\mathbf{J}^2/4n_s e^2 = m\mathbf{J}^2/2n_s e^2$. Using Ampère's law ($\boldsymbol{\nabla}\times\mathbf{B} = \mu_0\mathbf{J}$), we obtain the variational energy for the London theory:

$$F_{\text{London}}[\mathbf{B}] = \frac{1}{2\mu_0}\int d^3\mathbf{r}\left[\lambda_L^2\,(\boldsymbol{\nabla}\times\mathbf{B})^2 + \mathbf{B}^2\right] \tag{2.10}$$

where λ_L is a new length scale known as the **London penetration depth**:

$$\lambda_L = \left(\frac{m^*}{2n_s e^2\mu_0}\right)^{1/2} = \left(\frac{m_e}{n_s e^2\mu_0}\right)^{1/2} \tag{2.11}$$

Note that this is an electromagnetic length scale, in addition to the coherence length ξ of the Cooper pairs. So, unlike the case of the neutral superfluid, the physics of the charged superfluid has two different characteristic lengths. The appearance of this new length scale is the origin of many of the differences between neutral and charged superfluids.

[9] This is the conventional definition of n_s.

We can minimise the above variational energy with respect to $\mathbf{B}$. At the minimum, if we change the field $\mathbf{B}$ by a small amount $\delta\mathbf{B}$, the London variational energy $F_{\rm London}[\mathbf{B}]$ should not have any variation to first order in $\delta\mathbf{B}$. This variation in F is:

$$\begin{aligned}\delta F_{\rm London} &\simeq \frac{1}{\mu_0}\int d^3\mathbf{r}\left[\lambda_{\rm L}^2\,(\boldsymbol{\nabla}\times\mathbf{B})\cdot(\nabla\times\delta\mathbf{B})+\mathbf{B}\cdot\delta\mathbf{B}\right]\\ &= \frac{1}{\mu_0}\int d^3\mathbf{r}\left[\lambda_{\rm L}^2\,\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{B})+\mathbf{B}\right]\cdot\delta\mathbf{B}\\ &= \frac{1}{\mu_0}\int d^3\mathbf{r}\left[-\lambda_{\rm L}^2\,\nabla^2\mathbf{B}+\mathbf{B}\right]\cdot\delta\mathbf{B}\end{aligned} \tag{2.12}$$

[**Exercise**: check these steps. Recall that $\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{a})=-\nabla^2\mathbf{a}+\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{a})$.] Since we require δF to vanish for any $\delta\mathbf{B}(\mathbf{r})$ (to first order in $\delta\mathbf{B}$), the integrand must vanish everywhere. So, we obtain the **London equation** for the magnetic field inside a superconductor:

$$\boxed{\lambda_{\rm L}^2\,\nabla^2\mathbf{B}=\mathbf{B}} \tag{2.13}$$

This equation shows that all components of the magnetic field must decay exponentially inside a superconductor. The length scale over which this decay occurs is the London penetration depth.

The London equation is often quoted for historical reasons in the following form

$$\mathbf{J}=-\mu_0\mathbf{A}/\lambda_{\rm L}^2=-(n_s e^2/m_e)\mathbf{A} \tag{2.14}$$

This form (2.14) only applies in one particular gauge for the vector potential: $\boldsymbol{\nabla}\cdot\mathbf{A}=0$. We know that this is the correct gauge because the continuity equation requires that $\boldsymbol{\nabla}\cdot\mathbf{J}=0$ for static configurations.

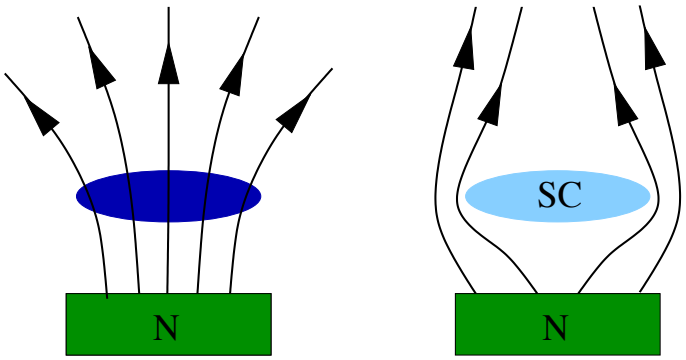


Figure 2.3: Meissner effect. The magnetic field penetrates a normal metal (left). Superconductors (SC) expel weak magnetic fields.

The London equation explains why a superconductor can expel magnetic fields. This is the **Meissner effect** and is the key signature for superconductivity.

This screening of the magnetic field in the interior of a Consider a semi-infinite superconducting slab with its surface parallel to the xy-plane. A uniform external field B_0 is applied parallel to this surface in the x-direction (Fig. 2.4). The London equation (2.13) tells us that the magnetic field $\mathbf{B}=(B_x,0,0)$ inside the superconductor obeys:

$$\lambda_{\rm L}^2\,\frac{d^2B_x}{dz^2}=B_x \quad\text{with}\quad B_x(z=0)=B_0 \tag{2.15}$$

where z is the depth from the surface of the superconductor into its bulk. This gives an exponential decay of the magnetic field into the bulk: $B_x(z)=B_0e^{-z/\lambda_{\rm L}}$ for $z>0$. So,

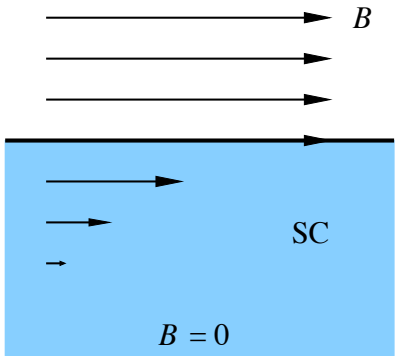


Figure 2.4: Meissner effect for semi-infinite superconductor slab. Magnetic field decays into the bulk. There is no field beyond a few penetration depths from the surface.

beyond a few penetration depths into the slab, the magnetic field is totally screened.In other word, the superconductor has expelled the magnetic field. superconductor is achieved by induced supercurrents which generate a magnetic field that cancels the applied field. In this geometry, Ampère's law says that

$$\mu_0 J_y = dB_x/dz = -(B_0/\lambda_{\rm L})\, e^{-z/\lambda_{\rm L}} \tag{2.16}$$

This induced current flows along a thin layer of thickness $\lambda_{\rm L}$ on the surface. The current per unit length is $I = \int_0^\infty J_y(z)\, dz = -B_0/\mu_0$. There is *no* current in the bulk of the superconductor. (Note that it does not matter whether I used the total B_x in Ampère's law, or just the induced field. The difference is the uniform applied field whose derivative is zero.)

3.2 Perfect conductor *vs* superconductor

The demonstration of a Meissner effect is a definitive signature of superconductivity. Indeed, it is a more important test of superconductivity than a vanishing electrical resistance. Other than the practical difficulties in distinguishing a small resistance from zero resistance, it is also a matter of principle that **a perfect conductor is not necessarily a superconductor**.

A perfect conductor or a superconductor can support **persistent currents**. For instance, once a current is set up in a ring, it will not decay. This is due to the lack of resistance to dissipate the energy in the current. A perfect conductor and a superconductor differ by how these persistent currents can be set up.

The Meissner effect expels a weak magnetic field. If we place the slab in the example above T_c in a magnetic field and then cool it down through the superconducting transiton, persistent screening currents will be generated on the surface as the temperature goes below T_C. The penetration depth becomes shorter and shorter as the temperature is lowered, so that more and more magnetic flux is expelled from the system.

However, the situation is quite different if we place a non-superconducting material in a magnetic field in a finite-resistance phase and then cool it down to a zero-resistance (but non-Meissner) state. In contrast to the superconductor, the system does *not* expel the magnetic field. In fact, a perfect conductor needs a time-varying magnetic field to generate screening currents *i.e.* Faraday's law. To be more precise, the ground state (or equilibrium state at $T > 0$) of a perfect conductor requires zero electric field in the bulk. Therefore, Faraday's law implies that $\boldsymbol{\nabla} \times \mathbf{E} = -\partial \mathbf{B}/\partial t = 0$. So, *the perfect conductor requires*

that there is no time variation in the magnetic field, while a superconductor prefers no magnetic fields at all.

In summary, the superconductor is something more than a metal with zero resistance. The Meissner effect is a macroscopic response of a superfluid condensate of Cooper pairs.

4 Flux quantisation

4.1 Flux quantum

We know in our study of neutral superfluids that vortices in the fluid have quantised circulation. An analogous phenomenon occurs for charged superfluids. Because a circulating electric current generates magnetic field, this phenomenon leads to the quantisation of the magnetic flux through a superconductor.

Consider a superconductor with a cylindrical cavity. A long thin solenoid runs through the centre of the cavity. When a current passes through the solenoid, a magnetic field B_0 exists inside the solenoid. The magnetic flux through a surface is the integral over the magnetic field through that surface: $\Phi = \int \mathbf{B} \cdot d\mathbf{S}$. So, the magnetic flux Φ_{app} due to the solenoid is simply the magnetic field B inside the solenoid multiplied by the cross-sectional area of the solenoid.

Note that the solenoid does not generate any magnetic field outside itself. If the electrons were classical particles, then they would not experience any Lorentz force at all. However, charged *quantum* particles can be physically affected by the vector potential alone, even in regions with zero magnetic field[10]. This is the Aharonov-Bohm effect. It tells us that the vector potential has *physical meaning*, and is not simply a mathematical way to express the zero-divergence magnetic field $\mathbf{B}$.

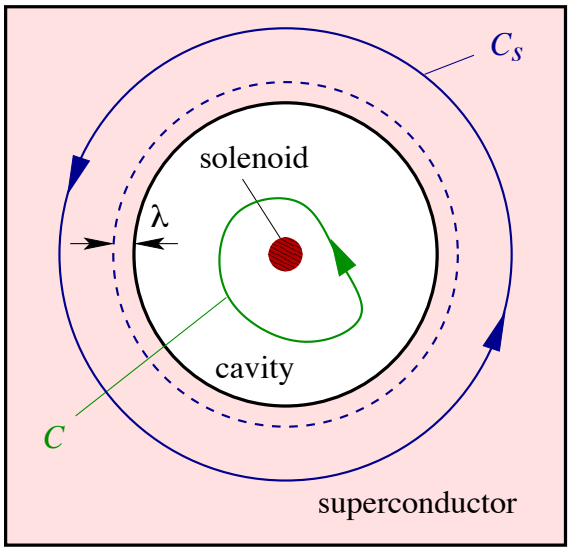


Figure 2.5: Superconductor with cylindrical cavity. A long thin solenoid introduces magnetic flux through the cavity. Screening currents are set up at the surface of the cavity (in the ring within the dashed line). The flux enclosed by C_s is quantised in units of $h/2e$.

The vector potential $\mathbf{A}_{\text{app}}$ outside the solenoid arising from the applied magnetic flux Φ_{app}

[10] ... in other words, even if there is zero probability of finding the particle in a region with non-zero magnetic field.

due to the solenoid is given by

$$\oint_C \mathbf{A}_{\rm app} \cdot d\mathbf{r} = \int \mathbf{B}_{\rm app} \cdot d\mathbf{S} = \Phi_{\rm app}$$

for any loop C inside the cavity (see Fig. 2.5) that encircles the solenoid. So, in a radially symmetric gauge, the vector potential circulates around the solenoid:

$$\mathbf{A}_{\rm app} = (\Phi_{\rm app}/2\pi r)\hat{\boldsymbol{\phi}} \tag{2.17}$$

where $\hat{\boldsymbol{\phi}}$ is the unit vector in the azimuthal direction in cylindrical polars.

To find the response of the superconductor to this magnetic flux, we should minimise the London variational energy (2.9):

$$F[\theta, \mathbf{A}] = \int d^3\mathbf{r} \left[\frac{n_s}{4m^*}(\hbar\boldsymbol{\nabla}\theta - 2e\mathbf{A})^2 + \frac{1}{2\mu_0}\mathbf{B}^2 \right] \tag{2.18}$$

We already know that we obtain the Meissner effect from this minimisation: for weak fields, there are no magnetic fields inside the superconductor (more than a few penetration depths away from any surface). This means that there are no induced currents inside the superconductor: $\boldsymbol{\nabla} \times \mathbf{B} = \mu_0\mathbf{J} = 0$. What does this mean for the phase of the order parameter $\psi_{\rm c}$? Recall that $\mathbf{J} \propto (\hbar\boldsymbol{\nabla}\theta - 2e\mathbf{A})$. So, we must have

$$\boldsymbol{\nabla}\theta = \frac{2e}{\hbar}\mathbf{A} \qquad \text{deep in superconductor} \tag{2.19}$$

In the context of the **Aharonov-Bohm effect**, we say that, as a Cooper pair moves from point 1 to 2, it picks up a phase change of

$$\Delta\theta_{12} = \int_1^2 \boldsymbol{\nabla}\theta \cdot d\mathbf{r} = \frac{2e}{\hbar}\int_1^2 \mathbf{A} \cdot d\mathbf{r}$$

If we take the particle around a closed loop C_s (see Fig. 2.5) in the superconducting ring, the phase change has to be an integer multiple of 2π. Therefore, we conclude that the magnetic flux $\Phi(C_s)$ encircled by the loop C_s must be:

$$\Phi(C_s) = \oint_{C_s} \mathbf{A} \cdot d\mathbf{r} = \frac{h}{2e}n_\Phi \qquad \text{integer } n_\Phi \tag{2.20}$$

Therefore we obtain **flux quantisation** in units of a *fundamental constant* the **flux quantum** Φ_0:

$$\boxed{\Phi_0 = h/2e} \tag{2.21}$$

Note that the factor of $2e$ for the flux quantum traces back to the charge of the Cooper pair. Therefore, the experimental verification of flux quantisation and the measurement of the flux quantum is a test of the existence of a superfluid of electron pairs.

You may have noticed that I have made a distinction between the flux $\Phi(C_s)$ through a loop C_s in the bulk of the superconductor and the applied flux $\Phi_{\rm app}$ from the solenoid. What if they are not the same? Something happens at the boundary of the cavity: the superconductor sets up circulating surface supercurrents which will induce a compensating magnetic field inside the cavity to make up for the difference between $\Phi(C_s)$ and $\Phi_{\rm app}$. The induced supercurrent changes the magnetic field distribution inside the cavity, changing the total flux to the nearest integer multiple of Φ_0.

In summary, we have found the magnetic flux enclosed by a multiply-connected superconductor is quantised in fundamental units of $h/2e$. It is a macroscopic manifestation of the Aharonov-Bohm effect.

4.2 Flux lines

What happens if we shrink the size of the cavity to zero in the example discussed above? This is analogous to our previous discussion of vortex lines in neutral superfluids: a line defect is formed in the order parameter ψ_c. We found that the order parameter ψ_c vanishes at the centre of this line defect, producing a normal core to the vortex line. The radius of the normal core is the coherence length ξ of the superconductor (2.5), analogous to the healing length for neutral superfluids.

If we have a normal core in the superconductor, it can be penetrated by an external magnetic field. This is analogous to the solenoid in the thought experiment depicted in Fig. 2.5. Then, the arguments in the previous section tells us that screening currents can be set up within a distance of the penetration depth from this line of magnetic flux such that there is no net magnetic field in the superconductor far from the line. In other words, for superconductors with $\lambda_L \gg \xi$, it is possible that *magnetic flux penetrates the system in vortex lines with quantised flux* Φ_0. This is analogous to vortex lines entering a rotating superfluid.

4.3 Two types of superconductors

We can now discuss the two ways in which the Meissner effect can break down. The Meissner effect of complete magnetic field expulsion only to *weak* magnetic fields.However, there must come a point when the field is so strong that it cannot be screened.

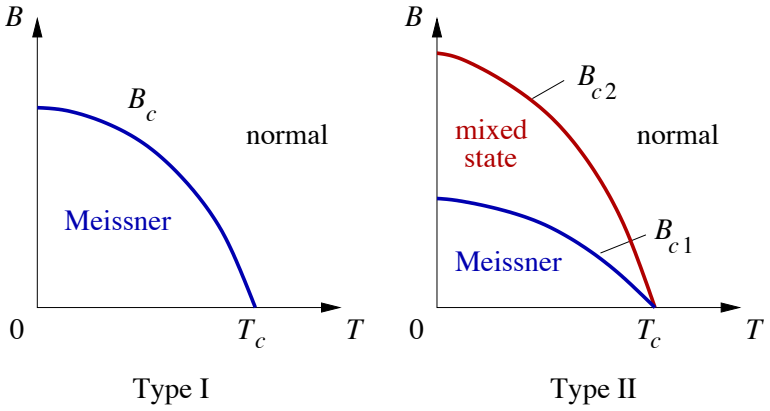


Figure 2.6: Two types of superconductors.

In a **type I superconductor**, the Meissner state gives way to the completely normal state where the magnetic field penetrates the sample completely. This occurs at a first-order phase transition as we raise the magnetic field beyond a critical value B_c where the superconducting condensation energy $f_{\rm cond}$ balances the energy cost in expelling the field $B^2/2\mu_0$. The measurement of the critical field in a type-I superconductor is therefore a measurement of the condensation energy: $f_{\rm cond} = B_c^2/2\mu_0$. In term of Ginzburg-Landau theory,

$$B_c = \frac{\Phi_0}{\sqrt{8}\pi\xi\lambda_L} \tag{2.22}$$

In contrast to the type-I superconductors, **type II superconductors** have two different critical fields. The Meissner effect of complete field expulsion is observed for field below

the lower critical field B_{c1}:

$$B_{c1} \simeq \frac{\Phi_0}{4\pi\lambda_{\rm L}^2} \ln\left(\frac{\lambda_{\rm L}}{\xi}\right) \qquad \text{for } \lambda_{\rm L} \gg \xi \tag{2.23}$$

At this point, quantised flux lines discussed in the previous section penetrate the sample. The stronger the field, the more flux lines there are. This is called the **mixed state**. The flux lines repel each other and form a triangular lattice. This is analogous to the state with quantised vortex lines penetrating a rotating neutral superfluid.

Superconducting order in a type-II superconductor is not destroyed until the field reach the upper critical field B_{c2}:

$$B_{c2} = \frac{\Phi_0}{2\pi\xi^2} \tag{2.24}$$

The flux lines are very densely packed as we reach B_{c2}. Since each flux line carries a flux of Φ_0 and B is the flux density through the sample, we see that the distance between the flux lines is of the order of ξ which is the normal core size. It is not surprising that the system is no longer a superfluid at B_{c2}.

How do we tell whether a superconductor is type I or II? The alert reader will have noticed that the picture of the flux line presented in the previous section makes sense only if the size of the normal core, which is the coherence length ξ, is smaller than the penetration depth $\lambda_{\rm L}$. More precisely, type-II behaviour is possible if the **Ginzburg-Landau parameter**: $\kappa = \lambda_{\rm L}/\xi$ is greater than $1/\sqrt{2}$. Otherwise, we have a type-I superconductor. This threshold can be determined by the condition that $B_{c2} > B_c$ so that, as we reduce the applied magnetic field from a high value, the mixed state becomes thermodynamically stable before the total flux expulsion becomes possible.

4.4 Comparison with neutral superfluids

We have already remarked that the flux penetration in a superconductor is analogous to the appearance of vortex lines in a superfluid under rotation. Indeed, the derivation of the critical fields in a superconductor follows similar arguments to the ones derived for the critical rotation frequencies in a neutral superfluid. For instance, both B_{c2} (2.24) for a superconductor and ω_{c2} (1.33) for a neutral superfluid have the same dependence on the coherence length ξ.

However, other results are qualitatively different. The presence of two characteristic length scales in superconductors gave us two critical fields B_{c1} (2.23) and B_{c2} (2.24). The neutral superfluid has only the healing length. The expression for ω_{c1} (1.32) uses the system radius as the second length scale. Consequently, a vortex line enters a macroscopically large system for nearly any angular velocity.

The key difference arises from the structures of the vortex line in a neutral superfluid and flux line in a superconductor is that the currents and fields around a flux line decays exponentially beyond a penetration depth from the line. This is the consequence of the London equation and is the same physics as the Meissner effect. The rest of the superconductor is essentially undisturbed by the presence of the flux line. Contrast this with a neutral vortex line where the tangential velocity decays as a power law $1/r$ so that its energy has a logarithmic divergence in the radial size R of the system. This is essentially because the neutral superfluid does not have a lengthscale analogous to the penetration depth, and

there is no cutoff until we reach the system size. If we regard the system size as replacing the penetration depth as a cutoff, we can say that the neutral superfluid is analogous to an extreme type-II superconductor ($\kappa \to \infty$).

5 Josephson Effect

We have learnt that the basic phenomena of superfluids and superconductors are connected to the phase θ of the condensate wavefunction ψ_c. However, the overall phase of an isolated system has no consequence for the behaviour of the system: a global phase change ($\theta(\mathbf{r}) \to \theta(\mathbf{r}) + C$) for the condensate wavefunction has no effect on any observable quantities. This means that the phase of the ground state is not observable. However, the ground state spontaneously breaks this global symmetry: the condensate does pick a definite phase in the ground state. How can this assertion be verified if the phase cannot be directly observed? To do so, we have to compare the phases of two different superconductors.

Suppose we have two superconductors both in their ground states. Each ground state will have its own overall phase so that they have condensate wavefunctions $\psi_c = |\psi_c| e^{i\theta_R}$ and $|\psi_c| e^{i\theta_L}$. Josephson pointed out in 1962 that, if we bring these two superconductors together, the *phase difference* $\theta_R - \theta_L$ can be measured as a current from one superconductor to the other. This current flows even if there is no voltage across the junction! This was experimentally verified one year later. This was arguably the final confirmation of BCS theory and of the central role that macroscopic phase coherence plays in superconductors. For extra reading, see Josephson's 1973 Nobel lecture.

5.1 DC Josephson current

Josephson considered two superconductors separated by a thin insulating layer. This is often called the superconductor-insulator-superconductor (SIS) junction. (The effect can also be generalised to the case of an intervening metallic layer.) An insulating layer presents a potential step to the electrons so that electron transport through the layer requires quantum-mechanical tunnelling.

Since these two superconductors are still separated by a barrier, the ground state of each superconductor are not strongly affected by the presence of the other superconductor. We can still regard the two systems as having independent condensates with phases θ_R and θ_L. Josephson predicted that, for an SIS junction, this phase difference alone causes a current across the junction, *without* any applied voltage across the junction.

To be more quantitative, let the superconductors on the left and right hand sides of the insulator have phases θ_L and θ_R respectively. In the absence of magnetic fields, the Josephson current would depend on the phase difference $\theta_R - \theta_L$ across the junction. In general, if there is a vector potential $\mathbf{A}$ in the system, then the current depends on a *gauge-invariant* version of this phase difference Θ_{LR} across the junction:

$$I_{L\to R} = I_c \sin \Theta_{LR} \qquad \text{with} \qquad \Theta_{LR} \equiv \theta_R - \theta_L - \frac{2e}{\hbar} \int_L^R \mathbf{A} \cdot d\mathbf{r} \tag{2.25}$$

where the line integral is taken along the short path that goes through the insulating gap from left to right. The insulating layer is usually very thin so that, in any sensibly chosen

gauge, the line integral is small and $\Theta_{\rm LR} \simeq \theta_{\rm R} - \theta_{\rm L}$. In principle, however, only the whole combination is gauge-invariant and has any physical significance. Here, we will treat the insulating layer as infinitessimally thin so that the line integral vanishes and that the phase has a discontinuity at the junction.

The microscopic mechanism for the Josephson current involves the coherent tunnelling of Cooper pairs instead of the tunnelling of single electrons. The information about the phase of the superconducting order parameter is carried by the Cooper pair from one superconductor to the other. The actual form of the Josephson current can be derived microscopically. However, I will simply justify it by noting that

- For a small phase difference $\Theta_{\rm LR}$, this Josephson pair-tunnelling current is approximately proportional to $\Theta_{\rm LR}$. This is analogous to what we know within each superconductor: gradients in the condensate phase are observable as superflow. More explicitly, $\Theta_{\rm LR} \simeq (\boldsymbol{\nabla}\theta - 2e\mathbf{A}/\hbar) \cdot \delta\mathbf{R}$ where $\delta\mathbf{R}$ is the short vector joining the two sides of the junction, and we recall that $\boldsymbol{\nabla}\theta - 2e\mathbf{A}/\hbar$ is proportional to the current inside a superconductor (2.8).
- For a larger phase difference, the sinusoidal nature of the relation (2.25) observes the fact that the phase is only defined[11] modulo 2π: there should be no physical consequence if $\theta_{\rm R}$ or $\theta_{\rm L}$ (and hence $\Theta_{\rm LR}$ is changed by an integer multiple of 2π.

Note that this Josephson current has a maximum value I_c. This is called the **critical current** of the junction. It depends on the transmission probability across the junction. For an insulating barrier, it should decrease exponentially if we increase the width of the junction. The tunnelling of a Cooper pair across the barrier involves breaking up and reformation of the Cooper pair and so I_c depends on the energy to break a Cooper pair. A microscopic calculation gives us the Ambegaokar-Baratoff formula:

$$I_c = \pi\Delta/2eR_n \tag{2.26}$$

where R_n is the resistance of the junction in the normal state (which is inversely proportional to the transmission probability) and Δ is half the binding energy of the Cooper pair.

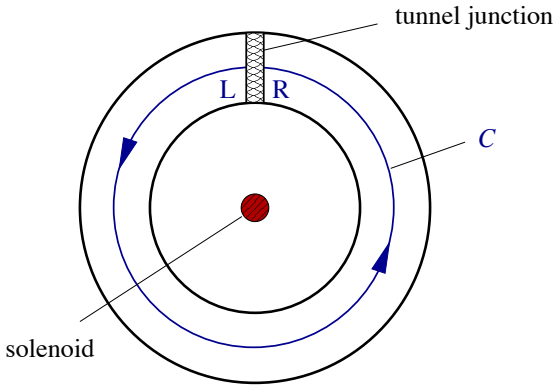


Figure 2.7: Superconducting ring with SIS junction. A solenoid through the centre provides magnetic flux Φ.

How do we set up this phase difference? One possible geometry is a split ring with the insulating layer inserted into the gap (Fig. 2.8). Suppose we apply a magnetic flux Φ

[11] In field theory, we say that the order parameter is compact.

through the centre of the ring. From the condition that there is no current inside the bulk of the superconductor, we know that $\boldsymbol{\nabla}\theta = 2e\mathbf{A}/\hbar$ [see (2.19)]. This means that the vector potential associated with the applied flux induces a phase twist inside the superconductor and so the superconducting condensate must have different phases on the two sides of the junction. The phase difference can be obtained by tracking how the phase changes along a path C which goes through the superconducting ring (rather than across the gap):

$$\theta_{\mathrm{R}} - \theta_{\mathrm{L}} = \int_C \boldsymbol{\nabla}\theta \cdot d\mathbf{r} = \frac{2e}{\hbar}\int_C \mathbf{A} \cdot d\mathbf{r} = \frac{2\pi\Phi}{\Phi_0}$$

Thus, there is a supercurrent of:

$$I = I_c \sin\Theta_{\mathrm{LR}} = I_c \sin\left(\frac{2\pi\Phi}{\Phi_0}\right) \tag{2.27}$$

across the insulating gap in the ring. This supercurrent varies sinusoidally with the applied flux Φ through the ring.

5.2 Interferometry for Cooper pairs

The SIS junction is the simplest demonstration of the dc Josephson effect. It forms the basis for 'superconducting quantum interference devices' — **SQUIDs**. A simple SQUID can be constructed by inserting two SIS junctions into a large superconducting ring (Fig. 2.8). This splits the ring into two halves. Each half of the ring is then attached to external leads so that we can pass a current I that enters the ring on one side of the junctions and exits the ring on the other. This current has two possible paths through the ring, each of which has to go through one of the junctions. (We should be careful to distinguish this transport current which transport charge into and out of the ring from any circulating currents in the SQUID.)

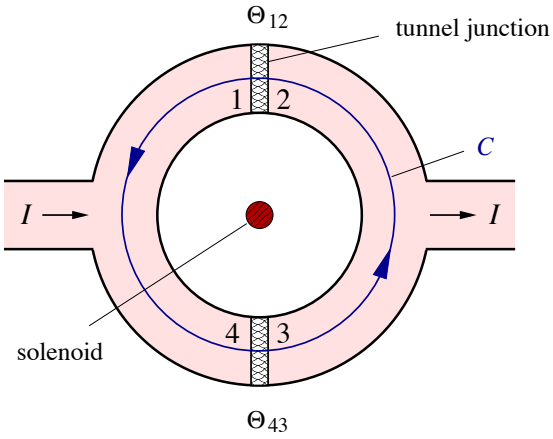


Figure 2.8: Schematic diagram of a SQUID. A solenoid through the centre provides magnetic flux Φ. A current I passes from left to right. Each of the two possible current paths involves Josephson tunnelling through a junction.

For simplicity, suppose that the two junctions have the same critical current I_c. Then, the total current I through the ring is the sum of the currents from each arm of the ring:

$$I = I_c(\sin\Theta_{12} + \sin\Theta_{43}) \tag{2.28}$$

where Θ_{12} and Θ_{43} are the gauge-invariant phase differences (from left to right in Fig. 2.8) across each of the two junctions.

In the absence of any flux through the ring, we expect the current through each arm to be the same. This determines the phase differences at the two junctions: $\Theta_{12} = \Theta_{43} = \sin^{-1}(I/2I_c)$.

If there is a flux of Φ through the ring, then the phase differences and the two currents will no longer be equal. The sum of the phase differences evaluated *clockwise* around the closed loop C in Fig. 2.8 is given by:

$$\Theta_{12} + \Theta_{34} = \Theta_{12} - \Theta_{43} = \frac{2e}{\hbar} \oint \mathbf{A} \cdot d\mathbf{r} = 2\pi \frac{\Phi}{\Phi_0} \tag{2.29}$$

where the line integral over the vector potential is evaluated along an *anticlockwise* closed loop C around the ring. The total current through the ring can be written as[12]:

$$I = 2I_c \sin\left(\frac{\Theta_{12} + \Theta_{43}}{2}\right) \cos\left(\frac{\pi\Phi}{\Phi_0}\right) \tag{2.30}$$

For a given applied current I at a given flux Φ, the phases Θ_{12} and Θ_{34} will adjust to satisfy the two equations (2.29) and (2.30).

The maximum possible current through the SQUID depends on the enclosed flux:

$$I_{\max} = 2I_c \cos\left(\frac{\pi\Phi}{\Phi_0}\right) \tag{2.31}$$

The key point here is that the currents in the two arms exhibit **interference**. This has a direct analogue with Young's slits in optics where two light rays from two slits interfere due to phase differences arising from different path lengths. Here, the interference arises from the different Aharonov-Bohm phases picked up as the Cooper pair goes round the SQUID in different directions (Fig. 2.9). The current has total destructive interference at $\Phi = \Phi_0(M+1/2)$ for any integer M. Note that this interferometry provides a very sensitive

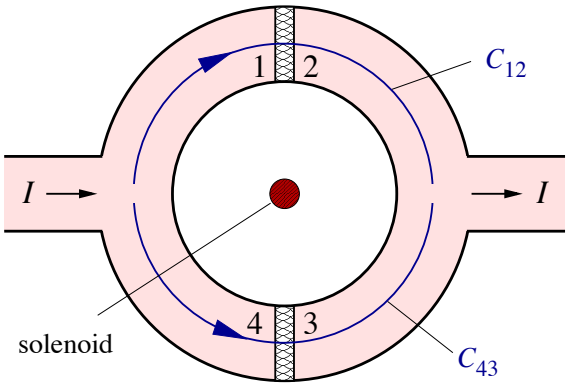


Figure 2.9: Aharonov-Bohm effect: interference between two current paths in a SQUID. The vector potential contributes a phase changes of $(2e/\hbar)\int_C \mathbf{A}\cdot d\mathbf{r}$ to the condensate wavefunction along the two paths $C = C_{12}$ and C_{43}. The difference in these two Aharonov-Bohm contributions is $2\pi\Phi/\Phi_0$. So, the interference is constructive when Φ/Φ_0 is an integer M and destructive when it is $M + 1/2$.

probe of changes in local magnetic fields. More recently, the combination of SQUIDs and small superconducting islands as capacitor-like elements has led to new developments in the field of superconducting electronics. In particular, many proposals have been put forward for quantum information processing. The advantage of using superconductors is the absence of decoherence effects in these systems.

[12] Recall that $\sin x + \sin y = 2[\sin(x+y)/2]\cos[(x-y)/2]$.

5.3 AC Josephson Effect

What happens if we do apply a dc voltage across an SIS junction? Josephson showed that this induces an *alternating* current! For an applied voltage V, the phase difference between the superconductors increases linearly in time, giving rise to an oscillating current:

$$\frac{d\Theta}{dt} = \frac{2eV}{\hbar} \quad \Rightarrow \quad I(t) = I_c \sin\left(\frac{2eV}{\hbar}\, t + \Theta(t=0)\right) \tag{2.32}$$

In terms of a microscopic picture, the tunnelling of the Cooper pair across the junction moves the pair to a region with lower chemical potential. The change in chemical potential is $\Delta\mu = -2eV$. So, we see that

$$\frac{d\Theta}{dt} = -\frac{1}{\hbar}\Delta\mu \tag{2.33}$$

This is called the **Josephson relation**.

We come to the conclusion that the connection between the chemical potential and the time evolution of the phase is quite fundamental. We can trace this back to basic time evolution of a quantum mechanical wavefunction. The wavefunction for a system at energy E evolves with a phase factor $e^{-iEt/\hbar}$. So, for two systems with an energy difference of $\Delta\mu$, this gives rise to a phase difference of $-t\,\Delta\mu/\hbar$. The significance of the ac Josephson effect is that this phase difference can be measured directly as a macroscopic current oscillation. The Josephson current is essentially $i(\Psi_{\rm R}^*\Psi_{\rm L} - \Psi_{\rm L}^*\Psi_{\rm R})/2$ where $\Psi_{\rm R,L}$ are the wavefunctions on either side of the SIS junction. Its existence demonstrates that we are right to attribute a complex order parameter with a global phase to superfluids and superconductors.

Chapter 3

Cooper Pairs

- Electron-phonon interaction
  - effective attraction between electrons near Fermi surface
  - real-space picture: retarded interaction
- Cooper instability for two electrons
  - electron pairing near Fermi surface for arbitrarily weak attraction
  - s-wave pairing: angular momentum $l = 0$, total spin $S = 0$
- Bardeen-Cooper-Schrieffer (BCS) theory
  - variational wavefunction for the ground state
  - number fluctuations
  - coherence factors
  - pairing symmetries: relation to Cooper pair spin state
  - excitation spectrum
  - Landau criterion revisited

In this chapter, we will discuss how an effective pairing interaction can give rise to Cooper pairs. This occurs even for a very weak attraction. We will then discuss the wavefunction proposed by Bardeen, Cooper and Schrieffer (BCS) for the ground state of a superconductor of Cooper pairs.

In simple metals, this effective pairing arises through electron-phonon interactions for electrons near the Fermi surface. This mechanism gives rise to superconductors with critical temperatures of the order of 30K. It is unlikely that a purely phonon mechanism will be able to explain the existence of the cuprate superconductors some of which remain superconducting at over 100K. A larger energy scale is needed: it is believed that the incipient antiferromagnetism of the cuprates provides this new energy scale. Although we do not understand the pairing mechanism in the cuprates, it appears that the superconducting state itself can be described by a wavefunction of the BCS type.

Neutral fermions can also pair up. This occurs in liquid ^{3}He and also in laser-cooled atoms. In fact, ultracold atoms can be tuned from repulsive to attractive in certain scattering channels using an external magnetic field (Feshbach resonances).

We will focus on the simplest Cooper pairing: s-wave pairing where the pairs have zero angular momentum and zero total spin. However, we should be aware that pairing in other angular-momentum channels are possible (*e.g.*, in liquid ^{3}He), giving rise to a rich variety of exotic phenomena.

1 Effective electron attraction

In normal metals, the long-ranged Coulomb interaction between bare electrons into a weak short-ranged interaction between quasiparticles. However, this is still a repulsive interaction. How can an attractive interaction arise? For conventional low-temperature superconductors, the answer lies in the interaction between the electrons and the ionic lattice.

We can see how an attractive interaction arises qualitatively in a real-space picture. A negatively charged electron would attract the positive ions in its vicinity, creating a net positive screening charge in the ionic charge distribution. (In second-quantised language, we would say that an electronic density perturbations create phonons.) The electron moves very fast: the Fermi velocity v_{F} in a metal can be a significant fraction of the speed of light while ions move at the speed of sound. This means that the lattice is still distorted long after the electron has gone. Therefore, *other* electrons are attracted towards this 'footprint' of the original electron. This creates an **indirect attraction** between different electrons which is not instantaneous but **retarded in time**. Moreover, since the original electron is now far away, this reduces its electrostatic repulsion with the electrons at the footprint.

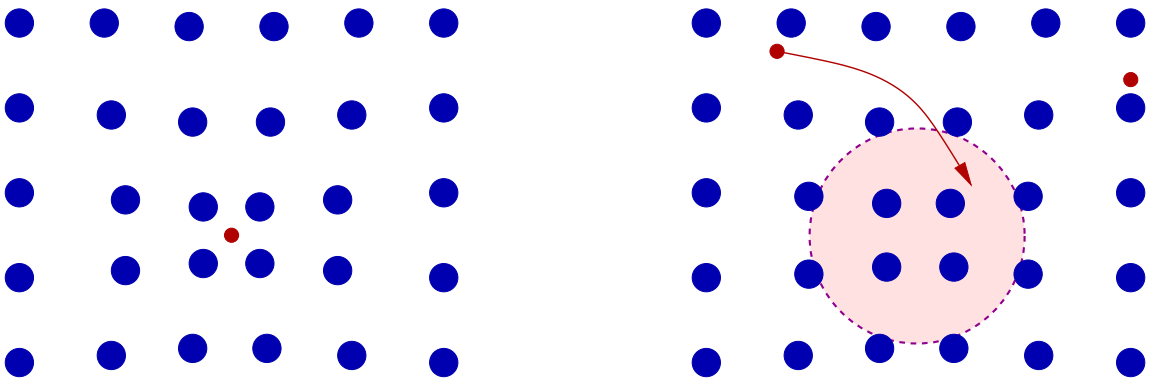


Figure 3.1: Electron leaves 'footprint' of positive ionic charge which attracts other electrons.

How does this affect the electronic states near the Fermi level? An electron density distribution $n(\mathbf{r}, t')$ at time t' affects the electrostatic potential energy $W(\mathbf{r}, t)$ for other electrons at a later time $t > t'$. For a small change in density, we can assume a linear response:

$$W(\mathbf{r}, t) = \int d^3\mathbf{r}' \int_{-\infty}^{t} dt'\, U(\mathbf{r} - \mathbf{r}', t - t')\, n(\mathbf{r}', t') \quad \Rightarrow \quad W_{\mathbf{q}\omega} = U_{\mathbf{q}\omega} n_{\mathbf{q}\omega} \tag{3.1}$$

where $U_{\mathbf{q}\omega} = \int d^3\mathbf{r} \int dt\, U(\mathbf{r}, t) e^{-i(\mathbf{q}\cdot\mathbf{r} - \omega t)}$ and similarly for the other Fourier amplitudes. An instantaneous interaction $U(t - t') \sim \delta(t - t')$ would give a frequency-independent $U_{\mathbf{q}\omega}$. A retarded interaction would have a frequency dependence. $U_{\mathbf{q}\omega}$ will be small at high frequencies at a fixed $\mathbf{q}$, because this corresponds to time intervals too short for a disturbance in one place to influence the potential at a distance $\sim 2\pi/q$ away from it.

We see that a periodic density variation $n_{\mathbf{q}}$ at wavevector $\mathbf{q}$ gives rise to a periodic potential energy landscape with the same periodicity. This will *diffract* for other electron waves,

changing their wavevectors by $\pm\mathbf{q}$. Moreover, the density oscillation at frequency ω creates an oscillating electric field for other electrons. From atomic physics, we know that this can excite an electron from its initial state to a final state which is higher in energy by $\hbar\omega$. This can be derived using time-dependent perturbation theory. These changes in momentum and energy for the other electrons must come from the electrons themselves. In other words, we can interpret the interaction as a pair of electrons *exchanging* momentum $\hbar\mathbf{q}$ and energy $\hbar\omega$ as they scatter off each other.

From the qualitative arguments at the start of this section, the electron-phonon interaction adds a retarded interaction between electrons. The exact form of $U_{\mathbf{q}\omega}$ depends on details of the phonon modes of the solid in question. One can construct a crude model to capture the essential physics by assuming that we have a homogeneous electron gas and point-like ions, all interacting through the Coulomb interaction. It can be shown that the effective interaction becomes:

$$U^{\text{eff}}_{\mathbf{q}\omega} = \frac{e^2}{\epsilon_0}\frac{1}{k^2_{\text{TF}}+q^2}\left(1+\frac{\omega^2_{\mathbf{q}}}{\omega^2-\omega^2_{\mathbf{q}}}\right) \tag{3.2}$$

where $\omega_{\mathbf{q}}$ is the frequency of a phonon with wavevector $\mathbf{q}$. The first term in U^{eff} comes from the direct electron-electron interaction with screening taken in account. The length scale over which this screening occurs is given by $1/k_{\text{TF}}$ with

$$k_{\text{TF}} = \left(\frac{e^2 g_{\text{F}}}{\epsilon_0}\right)^{1/2} = 0.8 k_{\text{F}}\left(\frac{r_s}{a_0}\right)^{1/2} \tag{3.3}$$

where g_{F} is the density of states at the Fermi level. This interaction is assumed to be instantaneous. This is justified in this context because the time scale for this mechanism is given by the plasma frequency ω_{p} of the metal, which is much higher than any other frequency scale in superconductors (for instance, $\hbar\omega_{\text{p}} \gg k_{\text{B}}T_c$).

The second term in U^{eff} comes from the electron-phonon interaction. This is obtained by considering the process where one electron emits a phonon which is absorbed by another electron. This is illustrated in Fig. 3.2. The key point is that *there is a range of frequencies*

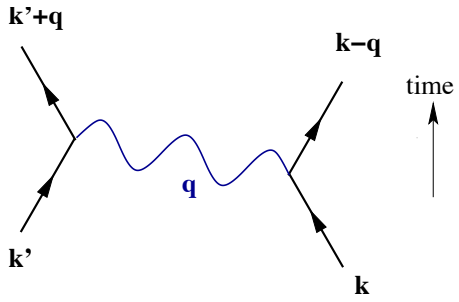


Figure 3.2: Electron interaction mediated by phonons. Phonon of momentum $\hbar\mathbf{q}$ emitted by electron with momentum $\hbar\mathbf{k}$ and absorbed by electron with momentum $\hbar\mathbf{k}'$.

below $\omega_{\mathbf{q}}$ where the effective interaction is attractive. This comes from the fact that the effective interaction mediated by phonons has the form $(\omega^2-\omega^2_{\mathbf{q}})^{-1}$. The divergence at $\omega=\omega_{\mathbf{q}}$ tells us that the coupling between the electrons and the lattice is resonant when the wavevector and frequency of the electron charge oscillations match the natural frequency of the phonons at the same wavevector. So, the transfer of energy between the electrons and the lattice is strongest near the resonance, leading to the divergence response $U^{\text{eff}}_{\mathbf{q}\omega}$. As with all driven resonant systems[1], at frequencies below the resonance, the ions are in

[1]Imagine a hanging spring used as a yoyo, driven by vertical oscillations of one end of the spring.

phase with the driving force in the sense that the ionic charge distribution has time to react to the electronic charge oscillations, more or less screening it. So, going back to our real-space picture, other electrons will be attracted towards the footprint of another electron when $\omega < \omega_{\mathbf{q}}$. On the other hand, at frequencies above the resonance, there is a π phase change to the response, and the ionic charge distribution is outo of phase with the electron charge oscillations. This means that the electrons are 'anti-screened'. In other words, there will be a net repulsion for $\omega > \omega_{\mathbf{q}}$. This is indeed what the above formula for U^{eff} indicates.

The effective interaction above is too complicated for a first attempt at the pairing mechanism. The phonons in solids form bands (just like electron waves). For the purposes of BCS theory, we will ignore the dispersion of the phonon frequencies, and simply identify a characteristic frequency ω_{D} for the phonon band. This is called the **Debye energy**. If we concentrate on physics at an energy scale much less than $\hbar\omega_{\text{D}}$, then we can simplify this further. This is justified because, for weak electron-phonon coupling, we will find that the binding energies of the Cooper pairs (and hence $k_{\text{B}}T_c$) are exponentially small compared to $\hbar\omega_{\text{D}}$. In this low-energy regime, the BCS model simply assumed that there is an **an attractive interaction for electrons with energy within a Debye energy from the Fermi surface**.

Since the electron energy $\varepsilon_{\mathbf{k}} \simeq E_{\text{F}} + \hbar v_{\text{F}}(k - k_{\text{F}})$ near the Fermi surface, this range of energies involves a thin shell of states in k-space with a range of wavevectors $\Delta k = \omega_{\text{D}}/v_{\text{F}}$. This gives us a characteristic length scale $1/\Delta k = v_{\text{F}}/\omega_{\text{D}}$ in the problem. It is the distance an electron has travelled by the time the disturbance caused by its presence has spread out. This is roughly the size of the smallest Cooper-pair wavepacket that can be constructed from momentum states within the restricted shell.

In the next section, we will see how this phonon-mediated attraction may give rise to the formation of electron pairs (Cooper pairs).

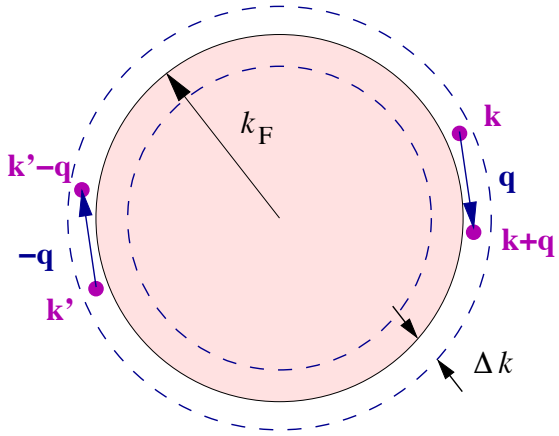


Figure 3.3: BCS interaction: attraction confined to thin shell of states above and below the Fermi surface. The energy range is the typical phonon energy $\hbar\omega_{\text{D}}$ (Debye energy). This corresponds a range of wavevectors $\Delta k = \omega_{\text{D}}/v_{\text{F}}$. Since $\hbar\omega_{\text{D}} \ll E_{\text{F}}$, the shell is very thin compared to the size of the Fermi surface k_{F}.

2 Cooper Instability

Before we study how N electrons behave in the BCS model, we can examine Cooper's even simpler version of the problem. He considered the effect of the attractive interaction on two electrons at the Fermi level (or just above), while all the states below the Fermi level are filled. In this treatment, all the electrons below the Fermi level do not interact.

They simply serve to prevent the two additional electrons from accessing the states below $E_{\rm F}$.

You may ask what happens if there are just 2 electrons. It can be shown that the two electrons in three dimensions would form a bound state if we have a sufficiently strong attractive interaction. As we will see below, the presence of the filled Fermi sea makes a dramatic difference: **two electrons at the Fermi sea will bind for an arbitrarily weak attractive interaction**. Discovering that this pairing occurs for two electrons inspired BCS to postulate a trial wavefunction which allows all the electrons to participate.

Returning to the Cooper problem, we can simply consider the two-particle wavefunction $\Psi(\mathbf{r}_1, \sigma_1, \mathbf{r}_2, \sigma_2)$ as long as we exclude all the momentum states below $k_{\rm F}$: $k_{1,2} > k_{\rm F}$. Since the interaction is isotropic for the spins, we should be able to factorise out a spin part for the wavefunction: $\Psi = \psi(\mathbf{r}_1, \mathbf{r}_2)\chi(\sigma_1, \sigma_2)$. There are 4 possible orthogonal spin states, spanned by the four possible choices for $|\sigma_1, \sigma_2\rangle$. It is best to use as a basis the four states classified by two quantum numbers: their total spin S and the z-component of the total spin S_z. For $S = 0$ (spin singlet), it is:

$$|S = 0\rangle = \left(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle\right)/\sqrt{2} \tag{3.4}$$

and the $S = 1$ triplet are:

$$\begin{aligned} |S = 1, S_z = 1\rangle &= |\uparrow\uparrow\rangle \\ |S = 1, S_z = 0\rangle &= \left(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle\right)/\sqrt{2} \\ |S = 1, S_z = -1\rangle &= |\downarrow\downarrow\rangle \end{aligned} \tag{3.5}$$

One reason why this choice of basis states makes physical sense is that these states have definite exchange symmetry: the singlet spin function is antisymmetric under exchange and the triplet spin functions are symmetric under exchange. In order that the overall exchange symmetry is antisymmetric for fermions, the spatial part $\psi(\mathbf{r}_1, \mathbf{r}_2)$ of the two-electron wavefunction must be symmetric for $S = 0$ and antisymmetric for $S = 1$. An antisymmetric $\psi(\mathbf{r}_1, \mathbf{r}_2) = -\psi(\mathbf{r}_2, \mathbf{r}_1)$ requires a node in the wavefunction. We expect the ground state to be nodeless, so we can focus on the $S = 0$ state with $\psi(\mathbf{r}_1, \mathbf{r}_2) = +\psi(\mathbf{r}_2, \mathbf{r}_1)$.

To obtain the spatial wavefunction, we take the interaction Hamiltonian for two particles which simply reduces to $\hat{H}_{\rm int} = U(\mathbf{r}_1 - \mathbf{r}_2)$. So, the spatial wavefunction obeys a Schrödinger equation:

$$\left[-\frac{\hbar^2}{2m}(\nabla_1^2 + \nabla_2^2) + U(\mathbf{r}_1 - \mathbf{r}_2)\right]\psi(\mathbf{r}_1, \mathbf{r}_2) = (2E_{\rm F} + E)\psi(\mathbf{r}_1, \mathbf{r}_2) \tag{3.6}$$

where E is the energy of the pair measured relative to the energy of two non-interacting electrons at the Fermi level. $U(\mathbf{r})$ is negative for an attractive interaction. We will adopt the BCS form which uses a momentum-independent $U_{\mathbf{q}}$ which corresponds to:

$$U(\mathbf{r}) = -\tilde{U}\,\delta(\mathbf{r}). \tag{3.7}$$

Recall that the two-particle wavefunction should be contain momentum states below the Fermi level and that the phonon-mediated attraction only applies to k_1 and k_2 within $\omega_{\rm D}/v_{\rm F}$ from the Fermi surface: $k_{\rm F} < k_{1,2} < k_{\rm F} + \omega_{\rm D}/v_{\rm F}$.

If the attractive interaction causes the two electron to form a bound state, then the energy should be lower than that of the two non-interacting electrons at the Fermi surface with

energy $2E_\mathrm{F}$. This means that E in the above equation will turn out to be negative and it will be the **binding energy** of the Cooper pair.

We can simplify the spatial wavefunction ψ. First of all, the interparticle interaction depends only on the relative distance between two particles, we can factorise out the motion of the centre of mass of the electron pair, $\mathbf{R} = (\mathbf{r}_1 + \mathbf{r}_2)/2$. In the absence of any external potential, the centre of mass should be in a momentum eigenstate, adding a kinetic energy $\hbar^2\mathbf{K}^2/4m$ to the energy of the system. So, the spatial wavefunction is $\psi(\mathbf{r}_1, \mathbf{r}_2) = e^{i\mathbf{K}\cdot\mathbf{R}}\phi(\mathbf{r}_1 - \mathbf{r}_2)/\sqrt{V}$. Substituting this form into the equation for ψ, we obtain a Schrödinger equation for ϕ which describes the relative motion:

$$\left[\frac{\hbar^2\mathbf{K}^2}{4m} - \frac{\hbar^2}{m}\nabla_\mathbf{r}^2 + U(\mathbf{r})\right]\phi(\mathbf{r}) = (2E_\mathrm{F} + E)\phi(\mathbf{r}) \tag{3.8}$$

For the ground state, we should choose $\mathbf{K} = 0$, *i.e.* the electron pair has zero momentum, and the momenta of the two electrons are equal and opposite: $\hbar\mathbf{k}_{1,2} = \pm\hbar\mathbf{q}$. So, we see that the restrictions on $\mathbf{k}_{1,2}$ also apply to $\mathbf{q}$. With this restriction in mind, the relative wavefunction ϕ can be written as:

$$\phi(\mathbf{r}) = \frac{1}{\sqrt{V}}{\sum_\mathbf{q}}' \phi_\mathbf{q} e^{i\mathbf{q}\cdot\mathbf{r}} \tag{3.9}$$

where the primed summation reminds us that the wavevectors should stay near the Fermi surface: $k_\mathrm{F} < q < k_\mathrm{F} + \omega_\mathrm{D}/v_\mathrm{F}$. In this problem, exchange symmetry means inversion symmetry: $\phi(\mathbf{r}_1 - \mathbf{r}_2) = \phi(\mathbf{r}_2 - \mathbf{r}_1)$; so, the Fourier components obey $\phi_\mathbf{q} = \phi_{-\mathbf{q}}$.

To deal with the wavevector restriction, we should work with the Schrödinger equation for $\phi(\mathbf{r})$ (3.8) in Fourier space. Taking the Fourier transform of the equation (3.8) with $\mathbf{K} = 0$ and $U(\mathbf{r}) = -\tilde{U}\delta(\mathbf{r})$, we obtain an equation for the Fourier amplitudes $\phi_\mathbf{k}$:

$$2\epsilon_\mathbf{k}\phi_\mathbf{k} - \frac{\tilde{U}}{V}{\sum_\mathbf{q}}'\phi_\mathbf{q} = E\phi_\mathbf{k} \tag{3.10}$$

where $\epsilon_\mathbf{k}$ is the free electron energy relative to the Fermi level: $\epsilon_\mathbf{k} = \hbar^2k^2/2m - E_\mathrm{F}$. To be more explicit, I have substituted the Fourier decomposition (3.9) into (3.8), then multiplied both sides of the equation by $e^{-i\mathbf{k}\cdot\mathbf{r}}$, and then integrated over all space. This picks out the Fourier amplitude at $\mathbf{q} = \mathbf{k}$ because of the fact that $\int d^3\mathbf{r}\, e^{i(\mathbf{q}-\mathbf{k})\cdot\mathbf{r}} = V\delta_{\mathbf{k},\mathbf{q}}$. This method of solving a differential equation in terms of Fourier amplitudes is very common. You should make sure you understand it.

Rearranging the above equation gives us:

$$\phi_\mathbf{k} = \frac{C\tilde{U}}{2\epsilon_\mathbf{k} - E} \qquad \text{with} \quad C = {\sum_\mathbf{k}}'\phi_\mathbf{k} \tag{3.11}$$

Summing both sides over $\mathbf{k}$ again gives us a self-consistent equation for C which can only be satisfied if:

$$\frac{\tilde{U}}{V}{\sum_\mathbf{k}}'\frac{1}{2\epsilon_\mathbf{k} - E} = 1 \tag{3.12}$$

This gives us an implicit equation for the binding energy E of the Cooper pair. To proceed further, we can approximate the sum as an integral over the energy range $0 < \epsilon_\mathbf{k} < \hbar\omega_\mathrm{D}$.

Since the Debye energy is small compared to the Fermi energy, we are only integrating over a narrow range of energy over which the density of states $g(\epsilon)$ should be well approximated by its value at the Fermi level $g_{\rm F}$. So, for any arbitrary function $f(\epsilon)$, we can write

$$\sum_{\mathbf{k}}{}' f(\epsilon_{\mathbf{k}}) \to V \int_0^{\hbar\omega_{\rm D}} d\epsilon\, g(\epsilon) f(\epsilon) \simeq V g_{\rm F} \int_0^{\hbar\omega_{\rm D}} d\epsilon\, f(\epsilon) \tag{3.13}$$

where ϵ is the energy as measured from the Fermi level and V is the volume of the system. So, our self-consistent equation becomes:

$$\tilde{U} g_{\rm F} \int_0^{\hbar\omega_{\rm D}} \frac{d\epsilon}{2\epsilon - E} = \frac{\tilde{U} g_{\rm F}}{2} \ln\left(1 - \frac{2\hbar\omega_{\rm D}}{E}\right) = 1 \tag{3.14}$$

We see that the pair energy E is indeed negative, indicating a bound state.

This theory is really only valid for weak electron-phonon interaction. The influence of this interaction on the electrons is measured by the **dimensionless coupling constant**:

$$\lambda_{\rm BCS} = \tilde{U} g_{\rm F} \tag{3.15}$$

and the weak-coupling regime is defined by $\lambda_{\rm BCS} \ll 1$. In this regime, the pair energy is

$$E = -2\hbar\omega_{\rm D}\, e^{-2/\lambda_{\rm BCS}} \tag{3.16}$$

This is much smaller than $\hbar\omega_{\rm D}$, justifying our original assumption that the energy scale for Cooper pair physics is much smaller than the typical phonon energies. We see that **this bound state exists for arbitrarily weak attraction**. Note also that the dependence on $\hbar\omega_{\rm D}$ arises from the energy width of the effective attractive interaction, rather than the strength of the interaction $\tilde{U}$ (which has in fact only very weak dependence on $\omega_{\rm D}$).

What is the spatial size of the bound state? It can be shown that the mean-square size $\langle r^2 \rangle$ is given by:

$$\langle r^2 \rangle = \int d^3\mathbf{r} |\phi(\mathbf{r})|^2 \mathbf{r}^2 = \frac{1}{V} \sum_{\mathbf{k}}{}' |\boldsymbol{\nabla}_{\mathbf{k}} \phi_{\mathbf{k}}|^2 = \frac{4\hbar^2 v_{\rm F}^2}{3E^2} \tag{3.17}$$

So, the size of the Cooper pair is $\langle r^2 \rangle^{1/2} = (2v_{\rm F}/\sqrt{3}\omega_{\rm D}) e^{2/\lambda_{\rm BCS}}$. Recall that $1/\Delta k = v_{\rm F}/\omega_{\rm D}$ is the characteristic length scale which can be interpreted roughly as the smallest Cooper size possible given the wavevector restrictions. We see that, at weak coupling where this theory is valid, the Cooper pair is very loosely bound compared to this scale. In other words, the states involved to form the Cooper pair involve a range of wavevectors much smaller than $\Delta k = \omega_{\rm D}/v_{\rm F}$: the pairing amplitude $\phi_{\mathbf{k}}$ is significant in a much narrower range of $(\Delta k)_{\rm Cooper} \sim (\omega_{\rm D}/v_{\rm F}) e^{-2/\lambda_{\rm BCS}} \ll \Delta k$.

We should also note that the function $e^{-2/\lambda_{\rm BCS}}$ has no Taylor expansion around $\lambda_{\rm BCS} = 0$. [**Exercise**: try it and see what happens.] This means that the binding energy and other properties of the system cannot be calculated in a perturbation theory in the coupling constant $\lambda_{\rm BCS}$ where the attraction is treated as a weak perturbation to the non-interacting electrons. This is because perturbation theory always produces results as a Taylor expansion in $\lambda_{\rm BCS}$. This mathematical point has physical signficance: it indicates that the superconducting ground state is *qualitatively different* from the Fermi liquid ground state.

In summary, Cooper's toy model indicates that **the Fermi liquid is unstable to the formation of Cooper pairs in the presence of a retarded attractive potential for the quasiparticles above the Fermi surface**. In the next section, we will discuss the BCS trial wavefunction for the superconducting ground state where there are *many* Cooper pairs.

3 BCS Ground State

3.1 Variational Ground State

In the previous section, we have discussed the simplified model of interacting electrons studied by Bardeen, Cooper and Schrieffer, where attractive interactions exist for electrons within a Debye energy near the Fermi surface (Fig. 3.3). By looking at two electrons above the Fermi level, we are led to expect that electron pairing occurs for all the electrons. BCS proposed a trial wavefunction for the ground state based on this result.

It is most conveniently written in terms of electron creation operators, rather than as a (fully antisymmetrised) many-body wavefunction in coordinate representation. Let us see how to write this for just one Cooper pair. It has the wavefunction:

$$\Psi(\mathbf{r}_1,\sigma_1,\mathbf{r}_2,\sigma_2) = \left[\frac{1}{V}{\sum_{\mathbf{k}}}' \phi_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{r}_1-\mathbf{r}_2)}\right] \frac{|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle}{\sqrt{2}} \tag{3.18}$$

We can translate the state into operator form:

$$\hat{c}^\dagger_\beta \hat{c}^\dagger_\alpha |\text{vac}\rangle \;\leftrightarrow\; \Psi_2(\mathbf{r}_1,\mathbf{r}_2) = [\phi_\alpha(\mathbf{r}_1)\phi_\beta(\mathbf{r}_2) - \phi_\alpha(\mathbf{r}_2)\phi_\beta(\mathbf{r}_2)]/\sqrt{2}$$

with $\phi_\alpha = V^{-1/2} e^{i\mathbf{k}\cdot\mathbf{r}}|\uparrow\rangle$ and $\phi_\beta \sim V^{-1/2} e^{-i\mathbf{k}\cdot\mathbf{r}}|\downarrow\rangle$. We obtain:

$$|\Psi_2\rangle = {\sum_{\mathbf{k}}}' \phi_{\mathbf{k}} \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow} |\text{vac}\rangle \tag{3.19}$$

Note that, although it is not written explicitly, $(\mathbf{k}\downarrow, -\mathbf{k}\uparrow)$ contributions are included because there is a corresponding $-\mathbf{k}$ term for each $\mathbf{k}$ term in the sum.

We have to extend this to make $N/2$ Cooper pairs for a system of N electrons. In the coordinate representation, we can construct something like

$$\Psi(\mathbf{r}_1,\ldots,\mathbf{r}_N) = \mathcal{A}[\Psi_2(\mathbf{r}_1,\mathbf{r}_2)\Psi_2(\mathbf{r}_3,\mathbf{r}_4)\ldots\Psi_2(\mathbf{r}_{N-1},\mathbf{r}_N)] \tag{3.20}$$

where $\mathcal{A}$ indicates that we have to include all permutations of pairs with appropriate sign changes to make sure that the antisymmetric fermion exchange symmetry is observed. It is easier to write this using fermion creation operators whose anticommutation relations encode this exchange symmetry implicitly. Extending (3.19) to $N/2$ pairs, we can write:

$$|\Psi_N\rangle = \prod_{i=1}^{N/2} \left[\sum_{\mathbf{k}} \phi_{\mathbf{k}} \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow}\right] |\text{vac}\rangle \tag{3.21}$$

We can see that, in this wavefunction, the electrons have been paired up two by two in **time-reversed states**, *i.e.* states with opposite momenta and spins. The summation over $\mathbf{k}$ is no longer restricted to be above the Fermi surface because we have to allow for *all* the electrons in the Fermi sea to have a degree of pairing. Nevertheless, from what we have learnt from the Cooper problem (discussion at the end of section 2), we expect a significant pairing amplitude $\phi_{\mathbf{k}}$ only for momenta close to the Fermi surface.

This is unfortunately still a very unwieldy wavefunction to work with. BCS proposed an alternative trial wavefunction: $C\prod_{\mathbf{k}}(1+\phi_{\mathbf{k}}\hat{c}^\dagger_{\mathbf{k}\uparrow}\hat{c}^\dagger_{-\mathbf{k}\downarrow})|\text{vac}\rangle$. We can see that, by multiplying

out the brackets, we obtain again a sum over states with Cooper pairs with pairing amplitudes $\phi_{\mathbf{k}}$. One odd thing is that states with different numbers of Cooper pairs have been superposed. We will see in the next section that, because only states with a narrow range of particle numbers dominates in this wavefunction, this should not give rise to unphysical consequences if we have a large number of electrons.

The conventional normalised form of the BCS wavefunction is:

$$\boxed{|\mathrm{BCS}\rangle = \prod_{\mathbf{k}} \left(u_{\mathbf{k}} + v_{\mathbf{k}} \hat{c}^{\dagger}_{\mathbf{k}\uparrow} \hat{c}^{\dagger}_{-\mathbf{k}\downarrow}\right) |\mathrm{vac}\rangle} \qquad \text{with} \quad |u_{\mathbf{k}}|^2 + |v_{\mathbf{k}}|^2 = 1 \tag{3.22}$$

where the product is over *all* wavevectors, not just those with $k < k_{\mathrm{F}}$. I have made the correspondence $v_{\mathbf{k}}/u_{\mathbf{k}} \leftrightarrow \phi_{\mathbf{k}}$. The coefficients $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$ are traditionally called **coherence factors**. For a system with time-reversal symmetry, the ground state wavefunction is real and we should expect the coherence factors to be real.

Note that this wavefunction reduces to the ground state $|\mathrm{FG}\rangle$ of a non-interacting Fermi gas if $u_{\mathbf{k}} = 0$ and $v_{\mathbf{k}} = 1$ for $k < k_{\mathrm{F}}$ but $u_{\mathbf{k}} = 1$ and $v_{\mathbf{k}} = 0$ for $k > k_{\mathrm{F}}$. For a general BCS state, the above wavefunction tells us that there is a probability of $|v_{\mathbf{k}}|^2$ to find an electron pair at $(\mathbf{k}\uparrow, -\mathbf{k}\downarrow)$ and a probability of $|u_{\mathbf{k}}|^2$ of not finding such a pair. So, we can find particles above the Fermi energy (and holes below the Fermi energy) at wavevectors where neither of the coherence factors is small. We will see later that the affected region is within a narrow range of energies from the Fermi level. The width of this energy range is given by the binding energy of the Cooper pair.

What are these coherence factors? It is tempting to use the pairing amplitudes calculated in Cooper's toy problem for the pairing amplitudes here in the full BCS wavefunction. Unfortunately, this does not work well because we did not allow the two electrons above the Fermi surface to interact with any electrons just below the Fermi surface. To deduce the coherence factors, we will use this BCS wavefunction as a *variational* wavefunction and find the coefficients $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$ which would minimise the variational energy $\langle \mathrm{BCS}|\hat{H}|\mathrm{BCS}\rangle$.

Before discussing the variational treatment of the BCS Hamiltonian, I will discuss one unusual feature of the BCS wavefunction. For a general form of the coherence factors, *the BCS wavefunction does not have a definite total particle number*! This can be seen explicitly if we multiply out the factors in $|\mathrm{BCS}\rangle$. For instance, if the product involves only 4 wavevectors, then the three factors multiply out to give a superposition of states with 0, 2, 4, 6 and 8 electrons. This is an artefact of the BCS trial wavefunction.

More generally, there is a probability of $v_{\mathbf{k}}^2$ for finding one $(\mathbf{k}\uparrow, -\mathbf{k}\downarrow)$ electron pair at wavevector $\mathbf{k}$ and a probability of $u_{\mathbf{k}}^2$ for finding none. So, we would expect the total number of particles to be

$$\langle \hat{N} \rangle = \sum_{\mathbf{k}} \langle \mathrm{vac}|(u_{\mathbf{k}}^2 \times 0 + v_{\mathbf{k}}^2 \times 2)|\mathrm{vac}\rangle = 2 \sum_{\mathbf{k}} v_{\mathbf{k}}^2 \tag{3.23}$$

Similarly, the variance in the total electron number is:

$$\mathrm{var}\, N = \langle \hat{N}^2 \rangle - \langle \hat{N} \rangle^2 = 4 \sum_{\mathbf{k}} u_{\mathbf{k}}^2 v_{\mathbf{k}}^2 \tag{3.24}$$

The summation over $\mathbf{k}$ gives a quantity proportional to the volume of the system and hence the number of particles $\langle \hat{N} \rangle$. In other words, the root-mean-square variations $(\mathrm{var}\, N)^{1/2}$ in the particle number is proportional to $\sqrt{N}$, and the fractional fluctuations in the particle

number scales as $1/\sqrt{N}$. This is the usual result from central limit theorem for classical probability distributions. The important point here is that **the fractional number fluctuations are small in a macroscopically large system**. This justifies the use of the BCS wavefunction even though we know that the eigenstates of the BCS Hamiltonian should also be states of definite particle number.

3.2 Variational approach

We will now derive how the coherence factors depend on the BCS Hamiltonian. The BCS Hamiltonian for the many-electron system is:

$$\hat{H}_{\rm BCS} = \sum_{\mathbf{k}\sigma} \epsilon_{\mathbf{k}} \hat{c}^\dagger_{\mathbf{k}\sigma} \hat{c}_{\mathbf{k}\sigma} + \frac{1}{2V} {\sum_{\substack{\mathbf{k}_1\mathbf{k}_2\mathbf{p}\\ \sigma_1\sigma_2}}}' U_{\mathbf{p}}\, \hat{c}^\dagger_{\mathbf{k}_1+\mathbf{p},\sigma_1} \hat{c}^\dagger_{\mathbf{k}_2-\mathbf{p},\sigma_2} \hat{c}_{\mathbf{k}_2,\sigma_2} \hat{c}_{\mathbf{k}_1,\sigma_1} \tag{3.25}$$

where V is the volume of the system, $\epsilon_{\mathbf{k}} = \hbar^2 k^2/2m - E_{\rm F}$ is the kinetic energy as measured from the Fermi level. The prime on the summation sign indicates that we should restrict the interaction term for states within $\hbar\omega_{\rm D}$ from the Fermi surface: $-\hbar\omega_{\rm D} < \epsilon_{\mathbf{k}} < \hbar\omega_{\rm D}$ for all the wavevector labels on the electron operators in the sum. The interaction $U_{\mathbf{p}}$ gives an attractive interaction if negative. We expect the interaction to have inversion symmetry $U(\mathbf{r}) = -U(-\mathbf{r})$ which means $U_{\mathbf{p}} = U_{-\mathbf{p}}$. We will specialise later to a momentum-independent $U_{\mathbf{p}} = -\tilde{U}$ which corresponds a point interaction in space: $U(\mathbf{r}-\mathbf{r}') \sim \delta(\mathbf{r}-\mathbf{r}')$.

Let us now use the variational principle to find the coherence factors $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$ and hence determine completely the BCS wavefunction (3.22). I will focus on the terms from the interaction that are only present in the BCS state but not the normal metal state. The variational energy is given by:

$$\langle {\rm BCS}|\hat{H}_{\rm BCS}|{\rm BCS}\rangle = 2\sum_{\mathbf{k}} \epsilon_{\mathbf{k}} v_{\mathbf{k}}^2 + \frac{1}{V}{\sum_{\mathbf{kq}}}' U_{\mathbf{k}-\mathbf{q}} u_{\mathbf{q}} v_{\mathbf{q}} u_{\mathbf{k}} v_{\mathbf{k}} \tag{3.26}$$

where $\epsilon_{\mathbf{k}} = \hbar^2 k^2/2m - E_{\rm F}$ and the primed summation restricts us to wavevectors where $-\hbar\omega_{\rm D} < \epsilon_{\mathbf{k}}, \epsilon_{\mathbf{q}} < \hbar\omega_{\rm D}$. The first term comes from the kinetic energy of the particles which have probability of $v_{\mathbf{k}}^2$ for being found at wavevector $\mathbf{k}$. I have absorbed into the Fermi energy term some contributions from the interaction which are second order in the creation and annihilation operators. These absorbed terms are also present in the metal and we will assume that they have not drastically altered the properties of the metal at low energies.

The second term in the variational energy drives superconductivity. Its effect is illustrated in Fig. 3.3. In particular, the important process comes from the scattering of electrons in spin-zero pairs of opposite momenta $\mathbf{k}_1 = -\mathbf{k}_2 = \mathbf{k}$ from $(\mathbf{k}\uparrow, -\mathbf{k}\downarrow)$ to $(\mathbf{q}\uparrow, -\mathbf{q}\downarrow)$ with a momentum transfer of $\hbar(\mathbf{k}-\mathbf{q})$. So, the initial state has probability amplitude $v_{\mathbf{k}} u_{\mathbf{q}}$ for having a Cooper pair at $(\mathbf{k}\uparrow, -\mathbf{k}\downarrow)$ but not $(\mathbf{q}\uparrow, -\mathbf{q}\downarrow)$, and the final state has amplitude $u_{\mathbf{k}} v_{\mathbf{q}}$ for $(\mathbf{q}\uparrow, -\mathbf{q}\downarrow)$ occupied but not $(\mathbf{k}\uparrow, -\mathbf{k}\downarrow)$. Combining these probability amplitudes, we see that this contribution to the variational energy is proportional to $u_{\mathbf{k}} v_{\mathbf{k}} u_{\mathbf{q}} v_{\mathbf{q}}$ as shown. (Note that we can no longer use classical probabilities for individual $\mathbf{k}$ state, because the term scatters from one momentum state to another.)

Note that this term is not present in a normal metal because energy conservation requires $2\epsilon_{\mathbf{k}} = 2\epsilon_{\mathbf{q}}$. This means that, in a Fermi liquid at $T = 0$, the initial and final states involved

are either both occupied or both unoccupied, and so scattering from one to the other is not possible. For a BCS state however, we can find particles above the Fermi energy (and holes below the Fermi energy) in regions where neither of the coherence factors is small.

The normalisation constraint on $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$ in (3.22) means that it is useful to put $u_{\mathbf{k}} = \sin(\Omega_{\mathbf{k}}/2)$ and $v_{\mathbf{k}} = \cos(\Omega_{\mathbf{k}}/2)$. Then, the variational energy becomes

$$\langle \mathrm{BCS}|\hat{H}_{\mathrm{BCS}}|\mathrm{BCS}\rangle = 2\sum_{\mathbf{k}} \epsilon_{\mathbf{k}} \cos^2 \frac{\Omega_{\mathbf{k}}}{2} + \frac{1}{4V}{\sum_{\mathbf{kq}}}' U_{\mathbf{k}-\mathbf{q}} \sin\Omega_{\mathbf{k}} \sin\Omega_{\mathbf{q}} \tag{3.27}$$

We want to choose the 'angles' $\Omega_{\mathbf{k}}$ to minimise this energy:

$$\begin{aligned} 0 = \frac{\partial\langle \hat{H}_{\mathrm{BCS}}\rangle}{\partial\Omega_{\mathbf{k}}} &= -\epsilon_{\mathbf{k}} \sin\Omega_{\mathbf{k}} + \frac{\cos\Omega_{\mathbf{k}}}{2V}{\sum_{\mathbf{q}}}' U_{\mathbf{k}-\mathbf{q}} \sin\Omega_{\mathbf{q}} \\ \Rightarrow \qquad \epsilon_{\mathbf{k}} \tan\Omega_{\mathbf{k}} &= \frac{1}{2V}{\sum_{\mathbf{q}}}' U_{\mathbf{k}-\mathbf{q}} \sin\Omega_{\mathbf{q}} \end{aligned} \tag{3.28}$$

I have used the fact that $U_{\mathbf{q}} = U_{-\mathbf{q}}$. We will see that it is convenient to define the **BCS gap function**:

$$\Delta_{\mathbf{k}} = -\frac{1}{V}{\sum_{\mathbf{q}}}' U_{\mathbf{k}-\mathbf{q}} u_{\mathbf{q}} v^*_{\mathbf{q}} \tag{3.29}$$

This quantity has the dimensions of energy. It will become clear later why this is called a gap. We can now express the angle parameters $\Omega_{\mathbf{k}}$ in terms of the gap function and the free-electron dispersion: $\tan\Omega_{\mathbf{k}} = -\Delta_{\mathbf{k}}/\epsilon_{\mathbf{k}}$, which means that $\sin^2\Omega_{\mathbf{k}} = (1+\cot^2\Omega_{\mathbf{k}})^{-1} = \Delta^2_{\mathbf{k}}/(\Delta^2_{\mathbf{k}} + \epsilon^2_{\mathbf{k}})$. Substituting for all the 'angles' in (3.28) above, we find that the gap function obeys the **self-consistent gap equation**:

$$\Delta_{\mathbf{k}} = -\frac{1}{2V}{\sum_{\mathbf{q}}}' \frac{U_{\mathbf{k}-\mathbf{q}}\Delta_{\mathbf{q}}}{(\Delta^2_{\mathbf{q}} + \epsilon^2_{\mathbf{q}})^{1/2}} \tag{3.30}$$

If we now specialise to $U_{\mathbf{k}} = -\tilde{U}$, then we find that Δ becomes independent of momentum too, obeying the equation:

$$1 = \frac{\tilde{U}}{2V}{\sum_{\mathbf{q}}}' \frac{1}{(\Delta^2 + \epsilon^2_{\mathbf{q}})^{1/2}} \tag{3.31}$$

This is analogous to the equation for the binding energy in the Cooper two-electron problem (3.12). In fact, this is called a gap because Δ is the energy gap for the lowest excitation of the BCS state. It can be shown that the gap function 2Δ is the energy to break a Cooper pair in a BCS superconductor.

Since the gap equation (3.31) has no dependence on the direction of $\mathbf{q}$, we can replace the sum over wavevectors to an integral over energies, as we did for the two-electron problem in the previous section:

$$1 = \frac{\tilde{U} g_{\mathrm{F}}}{2} \int_{-\hbar\omega_{\mathrm{D}}}^{\hbar\omega_{\mathrm{D}}} \frac{d\epsilon}{(\Delta^2 + \epsilon^2)^{1/2}} = \lambda_{\mathrm{BCS}}\ \sinh^{-1}\left(\frac{\hbar\omega_{\mathrm{D}}}{\Delta}\right) \tag{3.32}$$

where $\lambda_{\mathrm{BCS}} = \tilde{U} g_{\mathrm{F}}$ is the BCS coupling constant, as previously defined in (3.15). In the weak coupling regime ($\lambda_{\mathrm{BCS}} \ll 1$), we find that the gap function is:

$$\boxed{\Delta = 2\hbar\omega_{\mathrm{D}}\, e^{-1/\lambda_{\mathrm{BCS}}}} \tag{3.33}$$

This is larger by a factor of $e^{1/\lambda_{\rm BCS}}$ than the binding energy (3.16) for the two-electron toy problem. The reason for the difference is that we have now included the interactions among states above *and below* the Fermi level.

It is instructive to examine the behaviour of the coherence factors. From the equation for $\Omega_{\mathbf{k}}$, we can easily obtain:

$$v_{\mathbf{k}}^2 = 1 - u_{\mathbf{k}}^2 = \frac{1}{2}\left[1 - \frac{\epsilon_{\mathbf{k}}}{(\epsilon_{\mathbf{k}}^2 + \Delta^2)^{1/2}}\right] \tag{3.34}$$

This is illustrated in Fig. 3.4. We can see that the difference with the case for non-interacting electrons occur mainly for electrons whose energy $\epsilon_{\mathbf{k}}$ (measured from $E_{\rm F}$) is of the order of Δ. The number of non-interacting electron states which lie within Δ from the Fermi level is approximately $V g_{\rm F} \times 2\Delta$. Assuming that these form pairs, then there are $N_{\rm pair} \sim V g_{\rm F} \Delta$ Cooper pairs. Note that the fraction of electrons affected by the pairing is of the order of $\Delta/E_{\rm F}$ which is very small for nearly all metals.

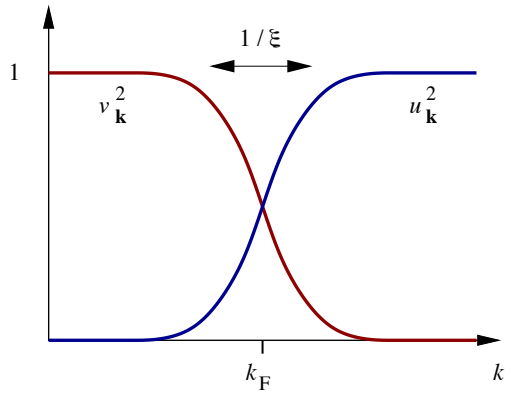


Figure 3.4: BCS coherence factors. Significant deviation from the Fermi gas form in the range of wavevectors $\Delta k = \Delta/v_{\rm F}$ near $k_{\rm F}$, corresponding to a range of $\pm\Delta$ in energy.

If we convert the energy window of $\pm\Delta$ to a wavevector scale, it would give an estimate of the size of the Cooper pair in this model, in the same way that we obtained a similar estimate for the Cooper problem for two electrons. Near the Fermi level, $\epsilon_{\mathbf{k}} \simeq v_{\rm F}(k - k_{\rm F})$. So, the energy range of Δ translate to a wavevector range of $\Delta/\hbar v_{\rm F}$. This gives us the **coherence length** of the superconductor:

$$\xi = \frac{\hbar v_{\rm F}}{\Delta} \tag{3.35}$$

3.3 Condensation energy

We can calculate the total energy of the ground state using the variational energy (3.26), and our result for the coherence factors. For a point interaction, it can be shown that the variational energy is

$$\langle \hat{H}_{\rm BCS} \rangle = 2 \sum_{k<k_{\rm F}} \epsilon_{\mathbf{k}} - \frac{g_{\rm F}\Delta^2}{2} V \tag{3.36}$$

The first term is simply the energy of a non-interacting Fermi gas. So, the onset of superconductivity has lowered the total energy. This can be regarded as the condensation energy of the system, which we have discussed in the context of Landau theory:

$$E_{\rm cond} = V g_{\rm F} \Delta^2/2 \,, \tag{3.37}$$

We know that the number of Cooper pairs is approximately $g_F V \times 2\Delta$. So, the condensation energy per Cooper pair is approximately $E_{\rm cond}/N_{\rm pair} \sim \Delta$ (up to factors of 2). This agrees with what we know from the Cooper instability of two electrons: the BCS gap as a binding energy for Cooper pairs.

3.4 Isotope effect

BCS theory depends on the electron-phonon interaction to give rise to an effective electron attraction. To test this hypothesis, we can modify the phonon properties of the system without changing the electronic properties by substituting different isotopes for the metals. It was found that there is indeed an **isotope effect** on the superconducting transition temperature T_c. If the isotope mass is changed by δM, the transition temperature changes by δT_c such that:

$$\frac{\delta T_c}{T_c} = -\beta \frac{\delta M}{M} \quad \text{with} \quad \beta \sim 0.5 \tag{3.38}$$

We will now see how the BCS model predicts $\beta \simeq 1/2$. In fact, the transition temperature is related to the BCS gap by a numerical factor: $\Delta(T=0) \simeq 1.7 k_B T_c$. So, we can simply see how Δ depends on the isotope mass.

From our expression for the gap Δ (3.33), we see that the energy prefactor is controlled by the Debye energy $\hbar\omega_D$. This is the main source of isotope mass dependence in this theory. The other source comes from the magnitude of the effective BCS interaction $\tilde{U}$.

To be see how the Debye energy depends on the mass, suppose we are considering acoustic phonons. The phonon frequencies should be lowered if we use a heavier isotope. In fact, ω_D should be inversely proportional to the mass M of the atom when we use different isotopes. This in turn means that the energy gap Δ should be proportional to $M^{-1/2}$. Therefore, for a small change in M, we expect $\delta\Delta/\Delta = -(\delta M/M)/2$. Given that $\Delta \propto T_c$, this implies that the above formula for the isotopic effect should have $\beta = 1/2$.

The experimental values for β are smaller than $1/2$. This has been attributed to the dependence of the effective interaction $\tilde{U}$ on the isotope mass. A detailed analysis of this dependence would require a more careful treatment of the balance between the screened Coulomb repulsion and the phonon-mediated attraction.

4 Gap function & pairing symmetry

The gap function $\Delta_{\mathbf{k}}$ is in fact a measure of the degree of pairing in the system and therefore can be regarded as the **order parameter for superconductivity**. To see this, consider first the *pairing amplitude* which measures the degree of pairing of electrons of opposite momenta and spin:

$$F_{\mathbf{k}} = \langle N+2 | \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow} | N \rangle \tag{3.39}$$

where $|N\rangle$ and $|N+2\rangle$ are the ground states with N and $N+2$ electrons respectively. As a simple example, we can calculate $F_{\mathbf{k}}$ for the single Cooper pair, $|\text{pair}\rangle = \phi_{\mathbf{k}} \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow} |\text{vac}\rangle$, in the Cooper problem, then we obtain

$$F_{\mathbf{k}} = \langle \text{pair} | \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow} | \text{vac} \rangle = \langle \text{vac} | \left(\phi^*_{\mathbf{k}} \hat{c}_{-\mathbf{k}\downarrow} \hat{c}_{\mathbf{k}\uparrow} \right) \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow} | \text{vac} \rangle = \phi^*_{\mathbf{k}} \tag{3.40}$$

Therefore, the pairing amplitude gives direct information of the pairing wavefunction $\phi(\mathbf{r})$.

For many particles, we can take advantage the BCS state which does not have a definite particle number, and simply evaluate the expectation value of the pair creation operator for the BCS state. This picks out the factor involving the pair at $\pm\mathbf{k}$:

$$\begin{aligned} F_{\mathbf{k}} &= \langle \mathrm{BCS}|\hat{c}^\dagger_{\mathbf{k}\uparrow}\hat{c}^\dagger_{-\mathbf{k}\downarrow}|\mathrm{BCS}\rangle = \langle \mathrm{vac}|\left(u^*_{\mathbf{k}} + v^*_{\mathbf{k}}\hat{c}_{-\mathbf{k}\downarrow}\hat{c}_{\mathbf{k}\uparrow}\right)\hat{c}^\dagger_{\mathbf{k}\uparrow}\hat{c}^\dagger_{-\mathbf{k}\downarrow}\left(u_{\mathbf{k}} + v_{\mathbf{k}}\hat{c}^\dagger_{\mathbf{k}\uparrow}\hat{c}^\dagger_{-\mathbf{k}\downarrow}\right)|\mathrm{vac}\rangle \\ &= u_{\mathbf{k}}v^*_{\mathbf{k}}\langle \mathrm{vac}|\hat{c}_{-\mathbf{k}\downarrow}\hat{c}_{\mathbf{k}\uparrow}\hat{c}^\dagger_{\mathbf{k}\uparrow}\hat{c}^\dagger_{-\mathbf{k}\downarrow}|\mathrm{vac}\rangle = u_{\mathbf{k}}v^*_{\mathbf{k}} \end{aligned} \tag{3.41}$$

To reach the second line, I have picked out the only term that has the same number of creation and annihilation operators. The last expectation value, it is just the squared magnitude of the state $\hat{c}^\dagger_{\mathbf{k}\uparrow}\hat{c}^\dagger_{-\mathbf{k}\downarrow}|\mathrm{vac}\rangle$ with one pair of electrons[2]. I have retained the complex conjugate on $v_{\mathbf{k}}$ to emphasize that this can be in general a complex quantity, but this does not happen for any ground state with time-reversal symmetry.

Comparing (3.41) with (3.29), we see that the **gap function** can be *defined* as:

$$\Delta_{\mathbf{k}} \equiv -\frac{1}{V}\sum_{\mathbf{q}} U_{\mathbf{k}-\mathbf{q}}\langle \hat{c}^\dagger_{\mathbf{q}\uparrow}\hat{c}^\dagger_{-\mathbf{q}\downarrow}\rangle \tag{3.42}$$

and the form given in (3.29) is just the form it takes for the BCS state at zero temperature. For instance, if we look at its temperature dependence, we will see that it decreases with increasing temperature, until it vanishes at a critical temperature T_c.

We have so far discussed an isotropic pairing wavefunction. In the Cooper problem, $\phi_{\mathbf{k}} \propto (E - 2\epsilon_{\mathbf{k}})^{-1}$ is independent of the direction of $\mathbf{k}$. In real space, this means that $\phi(\mathbf{r})$ depends only on the magnitude of $\mathbf{r}$ and not its direction. Similarly, in the BCS problem, we have obtained an isotropic gap function $\Delta_{\mathbf{k}} = \Delta$. This has the same symmetry as $l = 0$ atomic orbitals, and so this form of pairing is called ***s*-wave pairing**. This is the only possible pairing for a point-like attractive potential (momentum-independent $\tilde{U}$).

If we have a momentum-dependent interaction $U_{\mathbf{k}-\mathbf{q}}$, the scattering amplitude from $(\mathbf{k}, -\mathbf{k})$ to $(\mathbf{q}, -\mathbf{q})$ depends on the angle between the initial and final wavevectors[3]. This is in fact necessary if the BCS theory is to incorporate explain pairing in channels with higher l.

For instance, the gap function $\Delta_{\mathbf{k}}$ can change sign once as the azimuthal angle of $\mathbf{k}$ goes round by 2π. This has the same symmetry as $l = 1$ atomic states and is called p-wave pairing. This is seen in liquid ^{3}He and the recently discovered Sr_2RuO_4. Gap functions and pair wavefunctions with odd angular momenta break parity: $\Delta_{\mathbf{k}} = -\Delta_{-\mathbf{k}}$. This means that the pair amplitude and hence the pair wavefunction $\phi(\mathbf{r}_1 - \mathbf{r}_2)$ are odd under exchange. By the argument above about overall exchange symmetry, these are $S = 1$ superfluids/superconductors: they are *ferromagnetic*. We also know of systems where the gap function $\Delta_{\mathbf{k}}$ has d-wave symmetry ($l = 2$). This is in fact found experimentally in the cuprate superconductors. These d-wave superconductors have even parity in the pairing wavefunction and so the Cooper paris have $S = 0$.

You might ask why a system would prefer higher angular momentum states. One reason is that the effective electron-electron interaction is repulsive at short distance but attractive

[2] To evaluate this mechanically, use the fermion anticommutation relations to move the annihilation operators to the right.

[3] Since $\mathbf{k}$ and $\mathbf{q}$ are restricted to a thin shell near the Fermi surface, we can assume that their magnitudes are constant and the scattering amplitude is no dependence on them.

at larger distances. The electrons would want to pair up to take advantage of the attractive interaction, but prefer to have a pairing wavefunction that has a node at the origin in order to avoid each other at short distances.

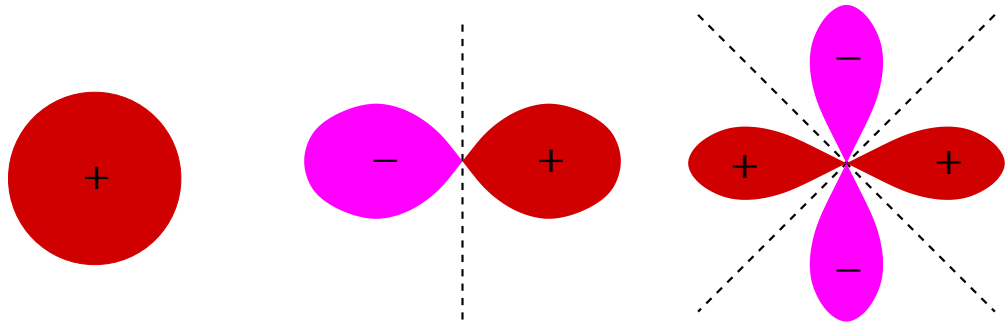


Figure 3.5: Pairing symmetries: the sign of the gap function $\Delta_{\mathbf{k}}$ in $\mathbf{k}$-space. Dashed lines indicate lines of nodes in the gap function. The total spin of the Cooper pair depends on the angular momentum to ensure overall fermionic exchange symmetry.

Further pairing symmetries are possible if the Fermi level passes through several bands. There will more than one Fermi surface and there could be pairing correlation between different Fermi surfaces. Such a scenario has been proposed for the iron pnictides which has several Fermi surfaces for different Fe d-bands. The parent compound is an antiferromagnetic metal with an ordering wavevector of (π,π). Interband scattering at a wavevector close to (π,π) introduces the possibility of the $s_{\pm}$ state where electron pairing is s-wave on each Fermi surface but the order parameter has opposite signs on two Fermi surfaces which are apart by (π,π) in k-space [4].

5 BCS Excitations

In this section, we will consider the fermionic excitations of a BCS superconductor. We will see that the gap Δ is the energy gap to the lowest excited state of the superconductor.

5.1 Excitations of a Fermi gas

Let us recall the excited states of a Fermi gas, before moving on to consider the superconducting system. The ground state of the Fermi gas has all the electron states occupied below the Fermi level. To make it look like the BCS wavefunction, we can write it as

$$|\mathrm{FG}\rangle = \prod_{k<k_\mathrm{F}} \hat{c}^\dagger_{\mathbf{k}\uparrow}\hat{c}^\dagger_{-\mathbf{k}\downarrow}\,|\mathrm{vac}\rangle$$

The basic fermionic excitations of the Fermi gas are simply the addition of electrons above the Fermi level, or the creation of holes below. The excited states are classified by

[4] I.I. Mazin, D.J. Singh, M.D. Johannes and M.H. Du, *Phys. Rev. Lett.*, **111** 057003 (2008).

momentum $\hbar\mathbf{k}$, spin σ and the energy $E_{\mathbf{k}}^{(0)}$ above the ground state[5]. If $k > k_{\rm F}$, then we are talking about adding an negatively-charged electron with momentum $\hbar\mathbf{k}$ and spin σ. If $k < k_{\rm F}$, then we are removing an electron of momentum $-\hbar\mathbf{k}$ and spin $-\sigma$ to make a positively-charged hole. We can denote the creation of an excitation by the operator $\hat{\gamma}^\dagger_{\mathbf{k}\sigma}$. Then, the excited states are:

$$\hat{\gamma}^\dagger_{\mathbf{k}\sigma}\,|\mathrm{FG}\rangle \quad \text{where} \quad \hat{\gamma}^\dagger_{\mathbf{k}\sigma} = \begin{cases} \hat{c}^\dagger_{\mathbf{k}\sigma} & \text{for } k > k_{\rm F} \\ \hat{c}_{-\mathbf{k},-\sigma} & \text{for } k < k_{\rm F} \end{cases} \quad \text{with energy:} \quad E_{\mathbf{k}}^{(0)} = |\epsilon_{\mathbf{k}}| \tag{3.43}$$

where $\epsilon_{\mathbf{k}} = \hbar^2k^2/2m - E_{\rm F}$.

You may be surprised that it costs energy to make a hole. After all, we are removing an electron with energy $\hbar^2k^2/2m$ so there will be less kinetic energy in the system. The point to remember is that we are working in the grand canonical ensemble. This means the system is free to exchange particles with a particle reservoir. The average number of particles in our system is controlled by the chemical potential $\mu = E_{\rm F}$. The minimisation principle that we need is that we have to minimise a *free* energy $\langle \hat{H} - \mu\hat{N}\rangle$. We have absorbed the chemical potential into the Hamiltonian for convenience. The contribution of the single-particle state at $\mathbf{k}$ to the revised Hamiltonian is $(\hbar^2k^2/2m - \mu)\hat{c}^\dagger_{\mathbf{k}}\hat{c}_{\mathbf{k}}$ (for each spin state). So, the addition of a particle at $\mathbf{k}$ increases the (free) energy, $\langle\hat{H}\rangle$, by $(\hbar^2k^2/2m - \mu)\times 1$. On the other hand, the removal of a particle at $-\mathbf{k}$ (with $k < k_{\rm F}$) also changes the (free) energy by $(\hbar^2k^2/2m - \mu)\times(-1)$ which is also positive. We can think of this as removing the electron at $-\mathbf{k}$ and then adding it to the particle reservoir which accepts particles at energy μ.

Another sense in which we can view the hole state as an excited state is that we can extract energy from it. A particle at $k_{\rm F}$ can drop from the Fermi level and fill up a hole. Thus, we can extract an energy of $E_{\rm F} - \hbar^2k^2/2m > 0$ from the system with a hole at $\mathbf{k}$ ($k < k_{\rm F}$). Put it another away, creating a hole in an N-electron gas makes an excited state of an $(N-1)$-electron gas.

We should verify several important properties for the creation and annihilation operators for excited states. These are trivial for the Fermi gas, but will be important for the superconductor. First of all, the ground state is, by definition, not excited: *the ground state is a vacuum of excitations* and should be destroyed by the annihilation operators for excitations. The annihilation operator is simply the Hermitian conjugate of the creation operator. So we have to check that:

$$\hat{\gamma}_{\mathbf{k}\sigma}\,|\mathrm{FG}\rangle = 0 \quad \text{where} \quad \hat{\gamma}_{\mathbf{k}\sigma} = \begin{cases} \hat{c}_{\mathbf{k}\sigma} & \text{for } k > k_{\rm F} \\ \hat{c}^\dagger_{-\mathbf{k},-\sigma} & \text{for } k < k_{\rm F} \end{cases} \tag{3.44}$$

This is clearly true for the Fermi gas because there are no electrons for $\hat{c}_{\mathbf{k}\sigma}$ to annihilate if $k > k_{\rm F}$, and Pauli exclusion prevents electron creation by $\hat{c}^\dagger_{-\mathbf{k},-\sigma}$ if $k < k_{\rm F}$. We also have to check that the excited state is *orthogonal* to the ground state:

$$\langle \mathrm{FG}|\,\hat{\gamma}^\dagger_{\mathbf{k}\sigma}\,|\mathrm{FG}\rangle = 0 \tag{3.45}$$

This is clearly true for the Fermi gas because the excited state has either one more or one fewer electron than the ground state.

[5]The (0) superscript is introduced to distinguish these excitation energies of a non-interacting Fermi gas from the excitation energies of the BCS state.

Finally, we can identify the statistics of these states. They have to be bosons or fermions. In this case, we know that they have half-integer spins and therefore must be fermions. Mathematically, we have to check that:

$$\{\,\hat{\gamma}_{\mathbf{k}\sigma}, \hat{\gamma}^\dagger_{\mathbf{k}'\sigma'}\,\} = \delta_{\mathbf{k},\mathbf{k}'}\,\delta_{\sigma,\sigma'}\,, \quad \{\,\hat{\gamma}_{\mathbf{k}\sigma}, \hat{\gamma}_{\mathbf{k}'\sigma'}\,\} = \{\,\hat{\gamma}^\dagger_{\mathbf{k}\sigma}, \hat{\gamma}^\dagger_{\mathbf{k}'\sigma'}\,\} = 0 \tag{3.46}$$

In other words, both electrons and holes are fermions. [**Exercise**: check this.] These anti-commutation relations also guarantee that these excited states are mutually orthonormal.

We can produce multiple excitations of this kind to produce excited states. Particularly important is the set of particle-hole excitations because they do not change the number of electrons. We have already discussed their role in the origin of intrinsic dissipation in Fermi liquids.

5.2 Excitations for the BCS state

We can now turn to the BCS state and find its elementary excitations. We know that the superconducting state affects only electrons within a BCS gap of the Fermi energy. In terms of wavevectors, these electrons have a range of wavevectors of about the inverse coherence length $1/\xi = \Delta/\hbar v_{\rm F}$. So, we would expect the excited state to be electron-like for $k \gg k_{\rm F}$ and hole-like for $k \ll k_{\rm F}$. Whereas the electron/hole character of the excitation switches over sharply at $k_{\rm F}$ for a Fermi gas, we might expect a smooth crossover, similar to the behaviour of the coherence factors $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$ (3.34).

Let us start with spin-up excitations. We can try a linear superposition of the form:

$$\hat{\gamma}^\dagger_{\mathbf{k}\uparrow} = A_{\mathbf{k}}\hat{c}^\dagger_{\mathbf{k}\uparrow} + B_{\mathbf{k}}\hat{c}_{-\mathbf{k}\downarrow}\,, \qquad \hat{\gamma}_{\mathbf{k}\uparrow} = A^*_{\mathbf{k}}\hat{c}_{\mathbf{k}\uparrow} + B^*_{\mathbf{k}}\hat{c}^\dagger_{-\mathbf{k}\downarrow} \tag{3.47}$$

Both terms in $\hat{\gamma}^\dagger_{\mathbf{k}\uparrow}$ causes a momentum change of $\hbar\mathbf{k}$ and adds spin $\Delta s_z = +1/2$ to the system. So, the state $\hat{\gamma}^\dagger_{\mathbf{k}\uparrow}|\text{BCS}\rangle$ has definite momentum and spin quantum numbers. This is sensible since the Hamiltonian conserves momentum and spin.

However, this does not respect number conservation: it is a strange particle/hole hybrid. This is a possible scenario because we have already broken number conservation in the BCS ground state. We should neverthless check that it can satisfy the basic properties of an excited state, as discussed in relation to the excitations of a Fermi gas. Since, they are built out of Fermi operators, they automatically satisfy Pauli exclusion: $(\hat{\gamma}^\dagger_{\mathbf{k}\uparrow})^2 = A_{\mathbf{k}}B_{\mathbf{k}}\{\,\hat{c}^\dagger_{\mathbf{k}\uparrow}, \hat{c}_{-\mathbf{k}\downarrow}\,\} = 0$. So, we see that they obey the second set of anticommutation relations in (3.46). For the first anticommutator to hold, we require

$$\{\,\hat{\gamma}_{\mathbf{k}\uparrow}, \hat{\gamma}^\dagger_{\mathbf{k}\uparrow}\,\} = |A_{\mathbf{k}}|^2\{\,\hat{c}_{\mathbf{k}\uparrow}, \hat{c}^\dagger_{\mathbf{k}\uparrow}\,\} + |B_{\mathbf{k}}|^2\{\,\hat{c}^\dagger_{-\mathbf{k}\downarrow}, \hat{c}_{-\mathbf{k}\downarrow}\,\} = |A_{\mathbf{k}}|^2 + |B_{\mathbf{k}}|^2 = 1 \tag{3.48}$$

(The anticommutor for $\hat{\gamma}_{\mathbf{k}\sigma}$ and $\hat{\gamma}^\dagger_{\mathbf{k}'\sigma'}$ for different wavevectors or spin is clearly zero, as required.)

We have also said that the ground state is a vacuum as far as the excitations are concerned. We need:

$$\hat{\gamma}_{\mathbf{k}\uparrow}\,|\text{BCS}\rangle = 0 \tag{3.49}$$

The BCS wavefunction contains many factors. The only one that is affected by this annihilation operator is the one involving a Cooper pair $(\mathbf{k}\uparrow, -\mathbf{k}\downarrow)$. The condition becomes

$$\hat{\gamma}_{\mathbf{k}\uparrow}(u_{\mathbf{k}} + v_{\mathbf{k}}\hat{c}^\dagger_{\mathbf{k}\uparrow}\hat{c}^\dagger_{-\mathbf{k}\downarrow})|\text{vac}\rangle \;\; = \;\; [A^*_{\mathbf{k}}v_{\mathbf{k}}\hat{c}_{\mathbf{k}\uparrow}\hat{c}^\dagger_{\mathbf{k}\uparrow}\hat{c}^\dagger_{-\mathbf{k}\downarrow} + B^*_{\mathbf{k}}u_{\mathbf{k}}c^\dagger_{-\mathbf{k}\downarrow}]|\text{vac}\rangle$$

$$\begin{aligned} &= [A^*_{\mathbf{k}} v_{\mathbf{k}} (1 - \hat{c}^\dagger_{\mathbf{k}\uparrow}\hat{c}_{\mathbf{k}\uparrow})\hat{c}^\dagger_{-\mathbf{k}\downarrow} + B^*_{\mathbf{k}} u_{\mathbf{k}} c^\dagger_{-\mathbf{k}\downarrow}]|\text{vac}\rangle \\ &= (A^*_{\mathbf{k}} v_{\mathbf{k}} + B^*_{\mathbf{k}} u_{\mathbf{k}}) c^\dagger_{-\mathbf{k}\downarrow}|\text{vac}\rangle = 0 \end{aligned} \tag{3.50}$$

So, we require $A^*_{\mathbf{k}} v_{\mathbf{k}} + B^*_{\mathbf{k}} u_{\mathbf{k}} = 0$. Combined with the previous condition (3.48), we see that $A_{\mathbf{k}} = u^*_{\mathbf{k}}$ and $B_{\mathbf{k}} = -v^*_{\mathbf{k}}$ will satisfy our the conditions.

We have now established the operator that creates a spin-up excitation with momentum $\hbar\mathbf{k}$. We can similarly perform the same procedure for spin-down states. We find that the excitations with momentum $\hbar\mathbf{k}$ and spin up or down above the BCS ground state are created by

$$\hat{\gamma}^\dagger_{\mathbf{k}\uparrow} = u^*_{\mathbf{k}}\hat{c}^\dagger_{\mathbf{k}\uparrow} - v^*_{\mathbf{k}}\hat{c}_{-\mathbf{k}\downarrow}, \quad \hat{\gamma}^\dagger_{\mathbf{k}\downarrow} = u^*_{\mathbf{k}}\hat{c}^\dagger_{\mathbf{k}\downarrow} + v^*_{\mathbf{k}}\hat{c}_{-\mathbf{k}\uparrow} \tag{3.51}$$

Note the different signs for the spin-up and spin-down operators. This is just a result of the way we defined the BCS wavefunction using $\hat{c}^\dagger_{\mathbf{k}\uparrow}\hat{c}^\dagger_{-\mathbf{k}\downarrow}$ rather than $\hat{c}^\dagger_{-\mathbf{k}\downarrow}\hat{c}^\dagger_{\mathbf{k}\uparrow}$ which is the same thing with the opposite sign.

Using the expressions (3.34) for the coherence factors (Fig. 3.4), we see that, indeed, these excitations are electron-like for wavevectors $\mathbf{k}$ far above the Fermi surface, and hole-like for $\mathbf{k}$ far below the Fermi surface.

Nature of the excitation

What is the nature of these excited states? In terms of exchange statistics, these excitations are fermionic: they are created by fermionic creation and annihilation operators. They are referred to as **quasiparticles** of a superconductor.

These BCS quasiparticles are quite different in nature to the quasiparticles of a normal Fermi liquid. The creation operator for the BCS quasiparticle looks quite strange. It seems to be part electron and part hole. For instance, at $k = k_{\text{F}}$, the excitation has probability amplitudes $(1/\sqrt{2})$ of creating an electron or creating a hole. (Note that this means that the excitations at k_{F} has no net charge.) We can say that this is just a consequence of the fact that our ground state wavefunction no longer has a definite particle number. It is perhaps not so surprising that our excitations are also do not respect number conservation.

It is however useful to consider in a bit more detail the action of the creation operator for an excitation, with, say, momentum $\hbar\mathbf{k}$ and spin $\uparrow$. For the excited state $\hat{\gamma}^\dagger_{\mathbf{k}\uparrow}|\text{BCS}\rangle$, the creation operator affects only the $(\mathbf{k}\uparrow, -\mathbf{k}\downarrow)$ pair. Assuming real coherent factors,

$$\begin{aligned} \hat{\gamma}^\dagger_{\mathbf{k}\uparrow}(u_{\mathbf{k}} + v_{\mathbf{k}}\hat{c}^\dagger_{\mathbf{k}\uparrow}\hat{c}^\dagger_{-\mathbf{k}\downarrow})|\text{vac}\rangle &= (u_{\mathbf{k}}\hat{c}^\dagger_{\mathbf{k}\uparrow} - v_{\mathbf{k}}\hat{c}_{-\mathbf{k}\downarrow})(u_{\mathbf{k}} + v_{\mathbf{k}}\hat{c}^\dagger_{\mathbf{k}\uparrow}\hat{c}^\dagger_{-\mathbf{k}\downarrow})|\text{vac}\rangle \\ &= (u^2_{\mathbf{k}}\hat{c}^\dagger_{\mathbf{k}\uparrow} - v^2_{\mathbf{k}}\hat{c}_{-\mathbf{k}\downarrow}\hat{c}^\dagger_{\mathbf{k}\uparrow}\hat{c}^\dagger_{-\mathbf{k}\downarrow})|\text{vac}\rangle \\ &= (u^2_{\mathbf{k}}\hat{c}^\dagger_{\mathbf{k}\uparrow} + v^2_{\mathbf{k}}\hat{c}_{-\mathbf{k}\downarrow}\hat{c}^\dagger_{-\mathbf{k}\downarrow}\hat{c}^\dagger_{\mathbf{k}\uparrow})|\text{vac}\rangle \\ &= [u^2_{\mathbf{k}}\hat{c}^\dagger_{\mathbf{k}\uparrow} + v^2_{\mathbf{k}}(1 - \hat{c}^\dagger_{-\mathbf{k}\downarrow}\hat{c}_{-\mathbf{k}\downarrow})\hat{c}^\dagger_{\mathbf{k}\uparrow}]|\text{vac}\rangle \\ &= (u^2_{\mathbf{k}} + v^2_{\mathbf{k}})\hat{c}^\dagger_{\mathbf{k}\uparrow}|\text{vac}\rangle = \hat{c}^\dagger_{\mathbf{k}\uparrow}|\text{vac}\rangle \end{aligned} \tag{3.52}$$

The second line tells us how $\hat{\gamma}^\dagger_{\mathbf{k}\uparrow}$ works. It inserts a spin-up electron if there is no electron pair (probability $u^2_{\mathbf{k}}$), or it removes a spin-down electron if there is a pair (probability $v^2_{\mathbf{k}}$). We can regard the latter process as removing the Cooper pair and then introducing the spin-up electron. In any case, the end result is that the excited state has a spin-up

electron at momentum $\hbar\mathbf{k}$ and no spin-down electron at $-\hbar\mathbf{k}$, instead of having a chance of $v_\mathbf{k}^2$ of having a $(\mathbf{k}\uparrow, -\mathbf{k}\downarrow)$ pair:

$$\hat{\gamma}_{\mathbf{k}\uparrow}^\dagger |\mathrm{BCS}\rangle = \hat{c}_{\mathbf{k}\uparrow}^\dagger \prod_{\mathbf{q}\neq\mathbf{k}} (u_\mathbf{q} + v_\mathbf{q} \hat{c}_{\mathbf{q}\uparrow}^\dagger \hat{c}_{-\mathbf{q}\downarrow}^\dagger)|\mathrm{vac}\rangle \tag{3.53}$$

Excitation energy

What is the energy of this excitation? We have added an electron at energy $\epsilon_\mathbf{k}$ if that state was not occupied (probability $|u_\mathbf{k}|^2$), or we would have removed energy $\epsilon_\mathbf{k}$ if the pair state was occupied (probability $|v_\mathbf{k}|^2$). So, assuming real coherence factors, the change in kinetic energy is $(u_\mathbf{k}^2 - v_\mathbf{k}^2)\epsilon_\mathbf{k}$. For the interaction energy (3.26) in the BCS ground state, we have only included Cooper pair interactions:

$$\langle \hat{H}_\mathrm{int} \rangle = -\frac{\tilde{U}}{V} {\sum_{\mathbf{kl}}}' u_\mathbf{l} v_\mathbf{l} u_\mathbf{q} v_\mathbf{q}$$

So, there is a change in interaction energy because the excited state has one Cooper pair missing: we lose the terms involving $u_\mathbf{k} v_\mathbf{k}$ in the expression above. This change is $+2(\tilde{U}/V) u_\mathbf{k} v_\mathbf{k} \sum_\mathbf{q} u_\mathbf{q} v_\mathbf{q} = +2\Delta u_\mathbf{k} v_\mathbf{k}$. There is a factor of 2 because either $\mathbf{q}$ or $\mathbf{l}$ can be $\mathbf{k}$. Using the form for the coherence factors (3.34), we find that the excitation energy is

$$E_\mathbf{k} = (u_\mathbf{k}^2 - v_\mathbf{k}^2)\epsilon_\mathbf{k} + 2\Delta u_\mathbf{k} v_\mathbf{k} = (\Delta^2 + \epsilon_\mathbf{k}^2)^{1/2} \tag{3.54}$$

We see that **the BCS gap is the lowest excitation energy of the superconductor**. There is an energy gap to excite these fermionic excitations. This should be contrasted with the absence of a gap in the analogous excitations (3.43) in a Fermi gas: $E_\mathbf{k}^{(0)} = |\epsilon_\mathbf{k}|$.

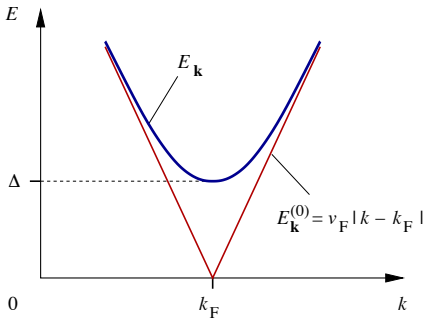


Figure 3.6: Excitation spectrum of the BCS state. $E_\mathbf{k}$ is the energy of a quasiparticle at momentum $\hbar\mathbf{k}$. The corresponding spectrum $E_\mathbf{k}^{(0)}$ for the Fermi gas is given for comparison.

I should point out that, although we have derived this for s-wave pairing, the expression for the excitation energy applies also for a non-isotropic order parameter:

$$\boxed{E_\mathbf{k} = (\Delta_\mathbf{k}^2 + \epsilon_\mathbf{k}^2)^{1/2}} \tag{3.55}$$

Since the order parameter $\Delta_\mathbf{k}$ has nodes for pairing in $l \neq 0$ channels, the excitation spectrum is gapless in specific directions in $\mathbf{k}$-space.

5.3 Superconductivity

We have now seen the basic elements of the BCS theory of superconductivity. We know an approximate wavefunction for the ground state as well as the excitations. How does

this give a microscopic explanation of the absence of electrical resistance and the Meissner effect? A proper treatment requires understanding how the BCS state responds to a small external vector potential. For resistivity, we need to consider a time varying potential to produce an electric field $\mathbf{E} = \partial\mathbf{A}/\partial t$ (in the gauge with the electric $\phi = 0$). In the case of the Meissner effect, we need to consider a static vector potential with a non-zero magnetic field $\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}$. This requires the calculation of the current correlation function of the system in zero field. Mathematical details can be found in Schrieffer's book[6]. I will simply discuss the results.

The key difference between a Fermi liquid and a superconductor is that there is an energy gap to creating excitations in superconductor. Whereas one can create particle-hole pairs at arbitrary low energy in a Fermi liquid, it costs an energy of 2Δ to break a Cooper pair and make two excited fermionic quasiparticles. This means that these fermionic excitations will not be produced by perturbations at frequencies less that $2\Delta/\hbar$. Then, the response of the superconductor to a static field is confined to the response of the condensate only, and this is described adequately by mean-field theories such as Ginzburg-Landau theory.

Indeed, the calculation for a Fermi liquid shows that the diamagnetic response is cancelled exactly by a paramagnetic response arising from particle-hole excitations. Once these excitations acquire an energy gap, this cancellation no longer holds and the system exhibits a Meissner effect.

A similar argument applies to conductivity. With a finite gap for a electron-like excitation, there are no final states for scattering by a (non-magnetic) impurity. As discussed at the very start of these lectures, this means that the fluid has no viscosity. Put another way, the excitation spectrum satisfies the Landau criterion (1.9) for superfluidity because there is a finite gap Δ: $E_{\mathbf{k}} - \mathbf{v} \cdot (\hbar\mathbf{k}) \simeq E_{\mathbf{k}} - \hbar v k_{\mathrm{F}} > \Delta - \hbar v k_{\mathrm{F}} > 0$ for small $\mathbf{v}$ provided that $\Delta > 0$.

I should point out that the application of the Landau criterion is not rigorous in the case of solid-state superconductors because the presence of an ionic lattice means that the electron fluid does not strictly obey Galilean invariance. Nevertheless, the simplified BCS theory is Galilean-invariant. So, the above argument should provide a qualitative explanation of the absence of resistance in an s-wave superconductor.

6 Summary

We have discussed the ground state of superconductors in the framework of the BCS theory. We have seen how an *arbitrarily weak attraction* interaction between electrons near the Fermi surface gives rise to an instability in a Fermi liquid to the formation of Cooper pairs. In fact, all we need is a finite density of states at the Fermi level. This theory has given us *microscopic* descriptions of several important quantities in the phenomenological Ginzburg-Landau theory of superconductors, such as the condensation energy and the coherence length. Most importantly, we have identified the BCS gap as an order parameter for superconductivity.

The theory also allows for an order parameter $\Delta_{\mathbf{k}}$ with different symmetries under spatial rotatation and inversion, giving rise to a variety of superconductors. The phenomenological framework can be adapted to describe these systems.

[6] J.R. Schrieffer, *Theory of Superconductivity*, Perseus Books.

The original BCS model is designed to study the limit of weak electron-phonon interactions. The attractive mechanism depends on a single dimensionless parameter $\lambda_{\rm BCS}$ (3.15). A more careful treatment is required to study the system at strong coupling ($\lambda_{\rm BCS} \sim 1$). We will have to consider how the interaction with phonons modifies the mass of the electron quasiparticles in the Fermi liquid. As we have already hinted in our discussion of the isotope effect, we also have study the balance between the screened Coulomb repulsion and the phonon-mediated attraction. It is found that the theory requires another parameter $\mu^* = g_{\rm F} U_{\rm Coulomb}$ to describe the Coulomb 'pseudopotential' for the electron quasiparticles. The discussion of the strong-coupling Eliashberg-McMillan theory and the derivation of the Coulomb pseudopotential is beyond the scope of this course.

Problem Sheet 1: Superfluids and Superconductors

1. Sound waves

This problem uses a different parametrisation of the fluctuations to obtain the collective waves of the Bose fluid from the Gross-Pitaevskii (GP) equation:

$$i\hbar\frac{\partial\psi_{\rm c}}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi_{\rm c} + u(|\psi_{\rm c}|^2 - \bar{n})\psi_{\rm c}$$

Consider only small fluctuations in the condensate wavefunction: $|\psi_{\rm c} - \sqrt{\bar{n}}| \ll \sqrt{\bar{n}}$. We can obtain a *linear* equation for the small perturbations by dropping higher-order terms.

(i) We can express the problem in natural lengths and time. The natural length is the healing length $\xi = (\hbar^2/mu\bar{n})^{1/2}$, and the natural time scale is derived from the Hartree energy $u\bar{n}$. With ξ as unit of length and $\hbar/u\bar{n}$ as unit of time, let us write

$$\psi_{\rm c}(\mathbf{r},t) = \sqrt{\bar{n}}\,[1 + f(\mathbf{s} = \mathbf{r}/\xi, \tau = u\bar{n}t/\hbar)]$$

Show that, for small fluctuations, the GP equation becomes:

$$i\frac{\partial f}{\partial\tau} = \left(-\frac{1}{2}\nabla^2 + 1\right) f + f^*$$

(ii) Try a wave solution: $f(\mathbf{s},\tau) = e^{i(\mathbf{q}\cdot\mathbf{s} - \Omega_{\mathbf{q}}\tau)}$. Why does it not work?

(iii) Using a solution which superposes time-reversed waves:

$$f(\mathbf{s},\tau) = u_{\mathbf{q}}e^{i(\mathbf{q}\cdot\mathbf{s} - \Omega_{\mathbf{q}}\tau)} + v_{\mathbf{q}}e^{-i(\mathbf{q}\cdot\mathbf{s} - \Omega_{\mathbf{q}}\tau)} \qquad (u, v \text{ real})$$

show that $\Omega_{\mathbf{q}}^2 = q^2(1 + q^2/4)$, and that, returning to dimensional quantities, the propagating waves with wavevector $\mathbf{k}$ have the frequency:

$$\omega_{\mathbf{k}} = \left[\frac{k^2}{2m}\left(2u\bar{n} + \frac{\hbar^2k^2}{2m}\right)\right]^{1/2}$$

2. London wavefunction

We cannot solve the problem of interacting bosons exactly. In the lectures, we have used a totally condensed state as an approximation to the ground state. It turns out that the concept of the condensate wavefunction remains useful for describing low-energy excitations for systems even when the interactions are strong. Suppose $\Psi_0(\mathbf{r}_1,\ldots,\mathbf{r}_N)$ is the exact ground state wavefunction. It has fast variations over interatomic distances because it encodes all the short-distance correlations between the repulsive bosons. London suggested that an approximate description of long-wavelength excitations of the system can be written as:

$$\Psi(\mathbf{r}_1,\ldots,\mathbf{r}_N) = \prod_{i=1}^{N}\left[f(\mathbf{r}_i)e^{i\theta(\mathbf{r}_i)}\right] \times \Psi_0(\mathbf{r}_1,\ldots,\mathbf{r}_N) \qquad (1)$$

where f and θ varies slowly over interatomic distances and $f \simeq 1$. For the totally condensed state, Ψ_0 is simply a constant and $\psi_{\rm c} = \sqrt{N}fe^{i\theta}$.

(i) Using the density operator $\hat{n}(\mathbf{r}) = \sum_i \delta(\mathbf{r} - \mathbf{r}_i)$, show that the the London wavefunction has a density profile:

$$n(\mathbf{r}) \equiv \langle \hat{n}(\mathbf{r}) \rangle = N f(\mathbf{r})^2 \int \left(\prod_{j \neq 1} d^3\mathbf{r}_j f(\mathbf{r}_j)^2 \right) |\Psi_0(\mathbf{r}, \mathbf{r}_2, \ldots)|^2$$

which is not simply $|\psi_{\rm c}|^2 = N f(\mathbf{r})^2$.

(ii) Using the current operator $\hat{\mathbf{J}} = (-i\hbar/2m) \sum_i [\delta(\mathbf{r} - \mathbf{r}_i)\boldsymbol{\nabla}_i + \boldsymbol{\nabla}_i \delta(\mathbf{r} - \mathbf{r}_i)]$, show that

$$\begin{aligned} \langle \hat{\mathbf{J}}(\mathbf{r}) \rangle &= \frac{\hbar}{m} \boldsymbol{\nabla}_{\mathbf{r}} \theta(\mathbf{r}) \times \int d^3\mathbf{r}_1 \ldots d^3\mathbf{r}_N \left(\prod_j f(\mathbf{r}_j)^2 \right) \sum_i \delta(\mathbf{r} - \mathbf{r}_i) |\Psi_0|^2 \\ &= \frac{\hbar}{m} n(\mathbf{r}) \boldsymbol{\nabla} \theta \end{aligned}$$

which shows that $\hbar \boldsymbol{\nabla} \theta / m$ is the superfluid velocity even in the case of a partially condensed state.

3. Vortex near a wall.

A vortex line with circulation K is at a distance d from a wall. The direction of line is parallel to the wall.

(i) What is the boundary condition for the velocity field $\mathbf{v}_s$ at the wall?

(ii) Show that the boundary condition can be satisfied by a suitable choice of an image vortex line.

(iii) Hence, show that the vortex line moves parallel to the wall (but transverse to the direction of the line) at a speed $v_L = K/4\pi d$.

4. Upper critical field of a type II superconductor

In this question, we will derive the upper critical field B_{c2} of a type II superconductor using Ginzburg-Landau theory. Therefore, we should consider minimising the Ginzburg-Landau free energy:

$$F_{\rm GL}[\psi_{\rm c}, \mathbf{A}] = \int d^3\mathbf{r} \left[\frac{1}{2m^*} |(-i\hbar\boldsymbol{\nabla} - 2e\mathbf{A})\psi_{\rm c}|^2 - \mu |\psi_{\rm c}|^2 + \frac{u}{2} |\psi_{\rm c}|^4 + \frac{1}{2\mu_0} \mathbf{B}^2 \right] \qquad (2)$$

with $\mu(T) = a(T_c - T)$ for some $a > 0$.

The order parameter $\psi_{\rm c}$ vanishes when the applied field B exceeds B_{c2}. So, just below B_{c2}, we expect the condensate wavefunction to be small and the system has a very weak Meissner response. (In terms of the flux lattice, the flux lines are tightly packed so that the normal cores nearly overlap. There is little condensate outside the cores so that there is little Meissner screening.) This means that we expect the magnetic field inside the system to be essentially the same as the applied field B.

With these approximations, the free energy $F_{\rm GL}$ is minimised by a *linear* Schrödinger equation for a condensate in a *uniform* magnetic field:

$$\frac{1}{2m^*}(-i\hbar\boldsymbol{\nabla} - 2e\mathbf{A})^2 \psi_{\rm c} = \mu\, \psi_{\rm c} \quad \text{with} \quad \mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}. \qquad (3)$$

(i) Take the direction of the applied field as the z-direction and use the Landau gauge $\mathbf{A} = (0, Bx, 0)$. Show that the solution of the Schrödinger equation can be written as $\psi_{\rm c}(\mathbf{r}) = e^{i(k_y y + k_z z)} f(x)$ where $f(x)$ satisfies

$$\frac{1}{2m^*}\left[-\hbar^2\frac{d^2}{dx^2} + (\hbar k_y - 2eBx)^2 + \hbar^2 k_z^2\right] f = \mu\, f \tag{4}$$

(ii) Why do k_y and k_z have to be real?

(iii) Compare equation (4) with the Schrödinger equation for a simple harmonic oscillator:

$$\left[-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{m\omega^2}{2}x^2\right]\psi = E\,\psi\,, \qquad E = \left(n + \frac{1}{2}\right)\hbar\omega \quad \text{for } n = 0, 1, 2, \ldots \tag{5}$$

Show that equation (4) has no solution if $e\hbar B/m^* > \mu$.

(iv) The coherence length is defined as $\xi = (\hbar^2/2m^* u|\psi_{\rm c}|^2)^{1/2}$ at zero applied field. Show that the upper critical field is given by:

$$B_{c2} = \frac{\Phi_0}{2\pi\xi^2} \qquad \text{with} \qquad \Phi_0 = \frac{h}{2e}.$$

5. Cooper instability at non-zero angular momentum

Consider the pairing instability of two electrons just above the Fermi surface in a two-dimensional system with area V. Denote the two-particle wavefunction as a product of the spin state χ and a spatial wavefunction ϕ for the relative coordinate $\mathbf{r}_1 - \mathbf{r}_2$:

$$\Psi(\mathbf{r}_1, \sigma_1, \mathbf{r}_2, \sigma_2) = \phi(\mathbf{r}_1 - \mathbf{r}_2)\chi(\sigma_1, \sigma_2) \quad \text{with} \quad \phi(\mathbf{r}_1 - \mathbf{r}_2) = \left[\frac{1}{V}{\sum_{\mathbf{k}}}' \phi_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{r}_1 - \mathbf{r}_2)}\right].$$

The prime over the summation sign indicates that the summation over $\mathbf{k}$ is restricted to the range: $k_{\rm F} < |\mathbf{k}| < k_{\rm F} + \omega_{\rm D}/v_{\rm F}$, where $v_{\rm F}$ is the Fermi velocity and $k_{\rm F}$ is the Fermi wavevector. In this question, we will discuss how the pairing amplitudes can be analysed in separate angular momentum channels.

(i) The pair wavefunction ϕ obeys the following Schrödinger equation:

$$-\frac{\hbar^2}{m}\nabla^2_{\mathbf{r}}\phi(\mathbf{r}) + U(\mathbf{r})\phi(\mathbf{r}) = (2E_{\rm F} + E)\phi(\mathbf{r})$$

where $U(\mathbf{r})$ is the interaction energy between two particles separated by $\mathbf{r}$, $E_{\rm F}$ is the Fermi energy and m is the mass of the electron. Show that the Fourier amplitude $\phi_{\mathbf{k}}$ of the above wavefunction obeys the equation:

$$(2\epsilon_{\mathbf{k}} - E)\phi_{\mathbf{k}} = -\frac{1}{V}{\sum_{\mathbf{q}}}' U_{\mathbf{k}-\mathbf{q}}\phi_{\mathbf{q}} \tag{6}$$

where $\epsilon_{\mathbf{k}} = \hbar^2 k^2/2m - E_{\rm F}$ and $U_{\mathbf{p}} = \int U(\mathbf{r}) e^{-i\mathbf{p}\cdot\mathbf{r}}\, d^2\mathbf{r}$.

(ii) The interaction $U_{\mathbf{k}-\mathbf{q}}$ scatters between pair states at $|\mathbf{k}, -\mathbf{k}\rangle$ and $|\mathbf{q}, -\mathbf{q}\rangle$. All the wavevectors in the above equations are very close to the Fermi surface. We can ignore the dependence of $U_{\mathbf{k}-\mathbf{q}}$ on the magnitudes k and q and assume that it is only a function of the angle $\theta_{\mathbf{k},\mathbf{q}}$ between the wavevectors $\mathbf{k}$ and $\mathbf{q}$: $U_{\mathbf{k}-\mathbf{q}} = U(\theta_{\mathbf{k},\mathbf{q}})$ (with the sign convention that it is the angle to rotate from $\mathbf{q}$ to $\mathbf{k}$). This must be a periodic function in θ. So, we can write:

$$U(\theta) = \sum_{m=-\infty}^{\infty} U_m e^{im\theta}$$

Let us also decompose the pairing amplitude into components with different angular dependence around the Fermi surface:

$$\phi_{\mathbf{k}} = \sum_{m=-\infty}^{\infty} \phi_m(k) e^{im\theta_{\mathbf{k}}}$$

where $\theta_{\mathbf{k}}$ is the angle between $\mathbf{k}$ and the positive x-direction, *i.e.* we are writing $\mathbf{k}$ in cylindrical polars. Show that the pairing amplitude for angular momentum m can be written as:

$$(2\epsilon_{\mathbf{k}} - E)\phi_m(k) = -U_m \int_{k_\mathrm{F}}^{k_\mathrm{F}+\omega_\mathrm{D}/v_\mathrm{F}} q\phi_m(q)\frac{dq}{2\pi}$$

(iii) Hence, show that a bound state exists ($E < 0$) with angular momentum m if U_m is negative and that the equation for the binding energy $|E|$ is

$$1 = -U_m \int_{k_\mathrm{F}}^{k_\mathrm{F}+\omega_\mathrm{D}/v_\mathrm{F}} \frac{k\,dk}{2\pi}\frac{1}{2\epsilon_{\mathbf{k}} + |E|} = -g_\mathrm{F} U_m \int_0^{\hbar\omega_\mathrm{D}} \frac{d\epsilon}{2\epsilon + |E|}$$

where g_F is the density of states at the Fermi level (per unit area) of the normal metal.

Remarks

It is possible that the interaction between fermions have $U_0 > 0$ preventing s-wave pairing. Then, the system will pair in the lowest-m channel with a negative U_m. This appears to be the case for the cuprate superconductors which have pairing in the $m = 2$ channel: 'd-wave' pairing. This means that the pairing amplitude changes sign 4 times as we go round the Fermi surface.

One way to construct this kind of interaction is to have a strong repulsion at short range while having an attraction if the particles are further apart at a range a, preferably with $1/a$ of the order of the Fermi wavevector.

6. BCS quasiparticles

Excitations with spin σ and energy $E_{\mathbf{k}} = (\Delta^2 + \epsilon_{\mathbf{k}}^2)^{1/2}$ are created in a BCS superconductor by the operators $\hat{\gamma}^\dagger_{\mathbf{k}\sigma}|\mathrm{BCS}\rangle$:

$$\hat{\gamma}^\dagger_{\mathbf{k}\uparrow} = u_{\mathbf{k}}\hat{c}^\dagger_{\mathbf{k}\uparrow} - v_{\mathbf{k}}\hat{c}_{-\mathbf{k}\downarrow} \qquad \text{and} \qquad \hat{\gamma}^\dagger_{-\mathbf{k}\downarrow} = u_{\mathbf{k}}\hat{c}^\dagger_{-\mathbf{k}\downarrow} + v_{\mathbf{k}}\hat{c}_{\mathbf{k}\uparrow}$$

where $\hat{c}^\dagger_{\mathbf{k}\sigma}$ creates an electron with wavevector $\mathbf{k}$ and spin σ and:

$$v_{\mathbf{k}}^2 = 1 - u_{\mathbf{k}}^2 = \frac{1}{2}\left[1 - \frac{\epsilon_{\mathbf{k}}}{(\Delta^2 + \epsilon_{\mathbf{k}}^2)^{1/2}}\right]$$

(i) $\hat{\gamma}^\dagger_{\mathbf{k}\uparrow}$ creates a fermion. Show that this means that we must have $u^2_{\mathbf{k}} + v^2_{\mathbf{k}} = 1$.

(ii) Using these coherence factors, describe whether the state $|\mathbf{k}\uparrow\rangle = \hat{\gamma}^\dagger_{\mathbf{k}\uparrow}|\mathrm{BCS}\rangle$ is electron-like or hole-like as a function of $\mathbf{k}$.

(iii) Find the probability of finding a spin-down electron with momentum $\hbar\mathbf{k}$ in the excited state $|\mathrm{exc}\rangle = \hat{\gamma}^\dagger_{\mathbf{k}\downarrow}|\mathrm{BCS}\rangle$. What is the probability of finding a Cooper pair of electrons in states $(\mathbf{k}\downarrow)$ and $(-\mathbf{k}\uparrow)$ in $|\mathrm{exc}\rangle$?

(iv) A superconductor does not absorb electromagnetic radiation at frequencies ω below $2\Delta/\hbar$. Explain.

MES

Mesoscopic Physics and Quantum Coherence

Edward McCann
Department of Physics, Lancaster University, Lancaster, LA1 4YB, United Kingdom

Mesoscopic physics is the name given to electronic behaviour in solid state nanostructures that are small so that their size is similar to relevant characteristic length scales. Examples of such length scales include the elastic mean free path (which governs the scale for ballistic transport), the phase coherence length (quantum interference effects), and the electronic wavelength (quantum confinement). The aim of this course is to describe key experimental transport phenomena including weak localisation, universal conductance fluctuations, Aharonov-Bohm oscillations, and conductance quantisation whilst giving an overview of theoretical methods such as the tight-binding model, the Landauer-Büttiker formalism, scattering theory, and scaling theory.

Contents

I. Introduction to Mesoscopic Physics and Quantum Coherence 1

II. The tight-binding model 2
A. General formalism 2
1. Lattice, crystal and basis 2
2. Tight-binding model 2
B. Example: monatomic square lattice 3
1. Lattice and tight-binding model 3
2. Expansion near the Γ point 3
C. Example: graphene 3
1. sp^2 hybridisation 3
2. The crystal structure of graphene 4
3. The reciprocal lattice of graphene 4
4. Calculation of the transfer integrals 4
5. Calculation of the energy 5
6. Expansion near the K points 5
7. Pseudospin and chirality 5
8. Kane-Mele spin-orbit coupling in graphene 6
D. Summary 7

III. Mesoscopic transport in the diffusive regime 7
A. Preliminaries 7
1. Periodic boundary conditions and density of states 7
2. Fermi function and electron density 8
3. Degenerate conductors 8
B. Characteristic scales 8
1. Fermi wavelength 8
2. Momentum relaxation 8
3. The Aharonov-Bohm effect 8
4. Phase relaxation 9
C. Classical regime 10
1. Classical equation of motion 10
2. Low field magnetoresistance 10
3. The Einstein model for degenerate conductors 11
D. Interference effects in mesoscopic electronics 11
1. Weak localisation 11
2. Altshuler-Aronov-Spivak (AAS) oscillations ($h/2e$) 13
3. Universal conductance fluctuations 13

IV. The Landauer-Büttiker formula 13
A. Conductance quantisation 13
B. The Landauer formula 14
C. Scattering theory in one dimension 15
1. Transfer matrix 16
2. Transfer matrix - example 16
3. Scattering matrix 16
4. Scattering matrix and current conservation 17
D. The Landauer-Büttiker formula 17

V. Localisation theory 17
A. Anderson localisation 18
1. Anderson tight-binding model of localisation 18
2. Fluctuations 19
3. Sub-critical disorder 19
B. Scaling theory 19
1. Thouless picture 19
2. The gang of four 20
3. Spin-orbit coupling 21

VI. Summary 22

References 22

I. INTRODUCTION TO MESOSCOPIC PHYSICS AND QUANTUM COHERENCE

Mesoscopic physics describes intermediate length scales between macro- and microscopic (typically 10 nm to 10 mm). The word 'mesoscopic' was coined by van Kampen: "The mesoscopic approach is more detailed than the macroscopic description, as it includes fluctuations, but has cut short the microscopic equations...." [N.G. van Kampen, Stochastic processes in physics and chemistry, p59, (North-Holland, Amsterdam 1981)]. These notes focus on the quantum mechanics and interference of electrons in small electronic devices at low temperature (phase-coherent devices), and they represent a very brief introduction to some concepts in mesoscopic physics. There are many excellent textbooks,

e.g., [1–6], and we refer the reader to them for further details and topics not covered here.

Following this brief introduction, Section II is a description of the tight-binding model with examples including the monatomic square lattice and graphene. Section III begins with some preliminary material about the two-dimensional electron gas (2DEG) such as density of states and carrier density, followed by the classical equation of motion for the drift velocity in a diffusive system leading to the dc conductivity and low-field magnetoresistance. Then, we describe some interference phenomena in disordered conductors, namely weak localisation and universal conductance fluctuations. The Landauer-Büttiker formula is introduced in Section IV, including its use within scattering theory to relate the conductance of a device to the probability that an electron can travel through it. Section V considers the case of strong disorder, and describes the scaling theory of localisation which provides a connection between the regimes of weak and strong localisation.

II. THE TIGHT-BINDING MODEL

The tight-binding model allows us to write down the single-particle Hamiltonian of a crystal structure analytically (i.e. using 'pen and paper') so it is a valuable theoretical tool. It can be written in real space and using second quantisation (often interaction effects can be added). Here, we'll assume translation invariance, use Bloch's Theorem, and derive the tight-binding Hamiltonian in wave vector space. First we'll introduce the model in general, then consider two examples in two dimensions.

A. General formalism

1. Lattice, crystal and basis

A Bravais lattice [7] is an infinite, regular array of discrete points such that all lattice points are equivalent, i.e. a (two-dimensional) Bravais lattice consists of all points with position vectors $\mathbf{R}$ of the form

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2, \tag{1}$$

where $\mathbf{a}_1$, $\mathbf{a}_2$ are primitive lattice vectors which span the lattice and n_1, n_2 take all integer values. Note that, for any given lattice, the choice of $\mathbf{a}_1$, $\mathbf{a}_2$ is not unique. Also, sometimes the word 'lattice' is used to describe crystals that are not a Bravais lattice, e.g. 'honeycomb lattice' (see example of graphene below). A primitive unit cell is an area of space that, when translated through all the vectors (1), just fills all space; it should contain one Bravais lattice point.

A crystal is not necessarily the same as a Bravais lattice. However, a crystal can be thought of as consisting of the Bravais lattice plus an arrangement of atoms with their constituent electronic orbitals located around each lattice point (this is known as the 'basis'), so crystal = Bravais lattice + basis.

2. Tight-binding model

Now we move onto the tight-binding model in wave vector space [6, 7]. Consider a system with n atomic orbitals ϕ_j per unit cell, labelled by index $j = 1 \ldots n$. It is assumed that the system has translational invariance. Then, the model may be written using n different Bloch functions $\Phi_j(\mathbf{k}, \mathbf{r})$ that depend on the position vector $\mathbf{r}$ and wave vector $\mathbf{k}$. They are given by

$$\Phi_j(\mathbf{k}, \mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \phi_j\left(\mathbf{r} - \mathbf{R} - \mathbf{s}_j\right) , \tag{2}$$

where the sum is over N different unit cells, $\mathbf{R}$ denotes the position vectors of the unit cells, and $\mathbf{s}_j$ denotes the position vector of the jth orbital with respect to the lattice vector of a given cell (i.e. with respect to the origin, the orbital is located at $\mathbf{R} + \mathbf{s}_j$).

We expand the electronic wavefunction as a linear expansion over the Bloch functions:

$$\Psi_{\mathbf{k}}(\mathbf{r}) = \sum_j \psi_j(\mathbf{k}) \Phi_j(\mathbf{k}, \mathbf{r}), \tag{3}$$

and substitute into the eigenvalue equation for the Hamiltonian $\mathcal{H}$:

$$\begin{aligned} \mathcal{H}\Psi_{\mathbf{k}}(\mathbf{r}) &= E\Psi_{\mathbf{k}}(\mathbf{r}), \\ \Rightarrow \qquad \sum_j \psi_j(\mathbf{k}) \mathcal{H} \Phi_j(\mathbf{k}, \mathbf{r}) &= E \sum_j \psi_j(\mathbf{k}) \Phi_j(\mathbf{k}, \mathbf{r}), \\ \sum_j \psi_j(\mathbf{k}) H_{ij} &= E \sum_j \psi_j(\mathbf{k}) S_{ij}, \end{aligned} \tag{4}$$

where we multiplied by $\Phi_i^*(\mathbf{k}, \mathbf{r})$ and integrated over all space. The transfer integral matrix elements H_{ij} and the overlap integral matrix elements S_{ij} are defined by

$$H_{ij} = \langle \Phi_i | \mathcal{H} | \Phi_j \rangle , \qquad S_{ij} = \langle \Phi_i | \Phi_j \rangle . \tag{5}$$

Eq. (4) can be written using $n \times n$ matrices as a generalised eigenvalue equation

$$H(\mathbf{k})\psi(\mathbf{k}) = E(\mathbf{k}) S(\mathbf{k}) \psi(\mathbf{k}) , \tag{6}$$

where ψ is an n component column vector, $\psi^T = (\psi_1, \psi_2, \ldots, \psi_n)$. The band energies E may be determined from the generalised eigenvalue equation (8) by solving the secular equation

$$\det\left(H - ES\right) = 0 , \tag{7}$$

where 'det' stands for the determinant of the matrix.

For simplicity, we will assume that the overlap integral matrix is a unit matrix, $S_{ij} = \delta_{ij}$: this is not a bad

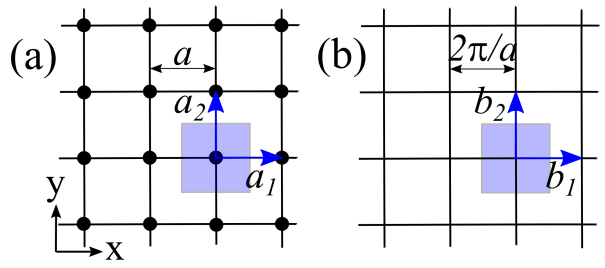


FIG. 1: (a) Monatomic square lattice and primitive unit cell (shaded square). (b) The reciprocal lattice (lattice points shown as dots) is a square Bravais lattice, the first Brillouin zone is also square (shaded).

approximation and it won't qualitatively change any of our conclusions (note that the calculation of S_{ij} proceeds in essentially the same way as H_{ij}, if you want to go through it yourself). Thus, we are left with a simple eigenvalue matrix equation:

$$H(\mathbf{k})\psi(\mathbf{k}) \;=\; E(\mathbf{k})\psi(\mathbf{k})\,. \tag{8}$$

By substituting the Bloch functions (2) into them, the transfer integral matrix elements (5) may be written as

$$\begin{aligned} H_{ij}(\mathbf{k}) \;&=\; \frac{1}{N}\sum_{\mathbf{R},\mathbf{R}'} e^{i\mathbf{k}\cdot(\mathbf{R}-\mathbf{R}')} \\ &\quad\times\langle\phi_i\left(\mathbf{r}-\mathbf{R}'-\mathbf{s}_i\right)|\mathcal{H}|\phi_j\left(\mathbf{r}-\mathbf{R}-\mathbf{s}_j\right)\rangle. \end{aligned}$$

This has a double sum over all lattice points. However, the matrix element on the second line only depends on $\mathbf{R}-\mathbf{R}'$, so one of the sums is redundant (imagine picking a single $\mathbf{R}'$ value, then doing the $\mathbf{R}$ sum. If you then move onto the next $\mathbf{R}'$ value, the $\mathbf{R}$ sum will be identical as all $\mathbf{R}'$ points are equivalent). We can choose $\mathbf{R}'=0$ and simplify the transfer integral matrix elements as

$$H_{ij}(\mathbf{k}) \;=\; \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} h_{ij}(\mathbf{R}), \tag{9}$$

$$h_{ij}(\mathbf{R}) \;=\; \langle\phi_i\left(\mathbf{r}-\mathbf{s}_i\right)|\mathcal{H}|\phi_j\left(\mathbf{r}-\mathbf{R}-\mathbf{s}_j\right)\rangle. \tag{10}$$

Now we will consider some examples of how to calculate the transfer integral matrix elements H_{ij} and determine the band structure.

B. Example: monatomic square lattice

1. *Lattice and tight-binding model*

We consider a monatomic square lattice with one s orbital per site, Fig. 1(a), the lattice constant is a. Primitive lattice vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ may be defined as

$$\mathbf{a}_1 = (a,0)\;, \qquad \mathbf{a}_2 = (0,a)\;, \tag{11}$$

where $a = |\mathbf{a}_1| = |\mathbf{a}_2|$. A primitive unit cell may be chosen as a square of side a.

The reciprocal lattice is a square Bravais lattice, Fig. 1(b), primitive reciprocal lattice vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ (we use $\mathbf{a}_1\cdot\mathbf{b}_1 = 2\pi$, $\mathbf{a}_1\cdot\mathbf{b}_2 = 0$, etc.) are given by

$$\mathbf{b}_1 = \left(\frac{2\pi}{a},0\right)\;, \qquad \mathbf{b}_2 = \left(0,\frac{2\pi}{a}\right)\;. \tag{12}$$

The first Brillouin zone is formed from perpendicular bisectors of the reciprocal lattice vectors, it's a square of side $2\pi/a$ containing wave vectors $\mathbf{k} = (k_x, k_y)$ with $-\pi/a < k_x \le \pi/a$ and $-\pi/a < k_y \le \pi/a$.

As there is only one orbital per unit cell, we can choose the atoms to sit on the lattice points (so $\mathbf{s}=0$) and the transfer integral gives the band energy,

$$\begin{aligned} E \equiv H(\mathbf{k}) \;&=\; \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} h(\mathbf{R}), \\ h(\mathbf{R}) \;&=\; \langle\phi\left(\mathbf{r}\right)|\mathcal{H}|\phi\left(\mathbf{r}-\mathbf{R}\right)\rangle. \end{aligned}$$

In the tight-binding model, we neglect contributions to the transfer integral that arise from interactions between atoms that are far apart. Here, we have just include an onsite energy ϵ_0 and four nearest-neighbour hoppings, each with the same parameter t:

$$\begin{aligned} \epsilon_0 \;&=\; h(0), \\ -t \;&=\; h(\mathbf{a}_1) = h(\mathbf{a}_2) = h(-\mathbf{a}_1) = h(-\mathbf{a}_2), \\ \Rightarrow \qquad E \;&=\; \epsilon_0 - t\left(e^{i\mathbf{k}\cdot\mathbf{a}_1} + e^{i\mathbf{k}\cdot\mathbf{a}_2} + e^{-i\mathbf{k}\cdot\mathbf{a}_1} + e^{-i\mathbf{k}\cdot\mathbf{a}_2}\right), \\ &=\; \epsilon_0 - 2t\left(\cos k_x a + \cos k_y a\right). \end{aligned}$$

The bandwidth $8t$ is set by the nearest-neighbour hopping parameter t, ϵ_0 just defines a reference level for measuring the energy from. Note that the tight-binding model doesn't tell us the value of the parameters ϵ_0 and t, they can perhaps be calculated using ab initio methods or determined by fitting to experimental data.

2. *Expansion near the* Γ *point*

The minimum of energy $E = \epsilon_0 - 4t$ occurs at the origin of reciprocal space $k_x = k_y = 0$, this point of symmetry is known as the Γ point. Near here, the dispersion is quadratic and isotropic. This can be seen by expanding the band energy (use $\cos x \approx 1 - x^2/2$ for $|x| \ll 1$):

$$E \;\approx\; \epsilon_0 - 4t + tk^2a^2 = \text{constant} + \frac{\hbar^2k^2}{2m}. \tag{13}$$

In other words, near the Γ point, the dispersion is that of free electrons with effective mass $m = \hbar^2/(2ta^2)$.

C. Example: graphene

1. *sp^2 hybridisation*

A carbon atoms has six electrons, two of them are core, four are valence electrons, the latter consist of one 2s and

three 2p orbitals [8]. In graphene there is sp^2 hybridisation: the single 2s and two of the 2p orbitals hybridise forming three 'σ bonds' in the x-y plane, while the remaining $2p_z$ orbital ('π' orbital) exists perpendicular to the x-y plane. When mixing with other atoms, this forms a π state which is extended across the sample. The π orbital is the only one relevant for energies of interest for transport measurements, so we only have to keep this one orbital per site in the tight binding model.

2. *The crystal structure of graphene*

The crystal structure of graphene [9–11] consists of carbon atoms arranged on a honeycomb structure, Figure 2(a). The honeycomb crystal structure of graphene consists of a hexagonal Bravais lattice with a basis of two atoms, labelled A and B, per unit cell. Primitive lattice vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ may be defined as

$$\mathbf{a}_1 = \left(\frac{a}{2}, \frac{\sqrt{3}a}{2}\right), \qquad \mathbf{a}_2 = \left(-\frac{a}{2}, \frac{\sqrt{3}a}{2}\right), \tag{14}$$

where $a = |\mathbf{a}_1| = |\mathbf{a}_2|$ is the lattice constant, the distance between adjacent unit cells, $a = 2.46\,\text{Å}$ [8]. Note that the carbon-carbon bond length is $a_{CC} = a/\sqrt{3} = 1.42\,\text{Å}$. Positions A and B are not equivalent; they can't be connected by a lattice vector.

3. *The reciprocal lattice of graphene*

The reciprocal lattice is a hexagonal Bravais lattice, Figure 2(b), primitive reciprocal lattice vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ are given by

$$\mathbf{b}_1 = \left(\frac{2\pi}{a}, \frac{2\pi}{\sqrt{3}a}\right), \qquad \mathbf{b}_2 = \left(-\frac{2\pi}{a}, \frac{2\pi}{\sqrt{3}a}\right). \tag{15}$$

The first Brillouin zone is a hexagon [shaded in Figure 2(b)], high-symmetry points are its centre (Γ), corners (K), and the mid-point of the side (M).

4. *Calculation of the transfer integrals*

There are two atoms per unit cell each with one p_z orbital. As all orbitals are identical, we will drop the j label on them, giving just $\phi(\mathbf{r})$. We relabel the basis index $j = 1, 2$ as $j = A, B$. To calculate the transfer matrix elements, we assume that the lattice points coincide with A atoms, $\mathbf{s}_A = 0$, so that the B atom in each unit cell is located at $\mathbf{s}_B = (0, a/\sqrt{3})$.

A diagonal transfer matrix element (9) is, e.g., given by

$$\begin{aligned} H_{AA}(\mathbf{k}) &= \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} h_{AA}(\mathbf{R}), \\ h_{AA}(\mathbf{R}) &= \langle \phi(\mathbf{r}) \,|\mathcal{H}|\, \phi(\mathbf{r}-\mathbf{R})\rangle. \end{aligned}$$

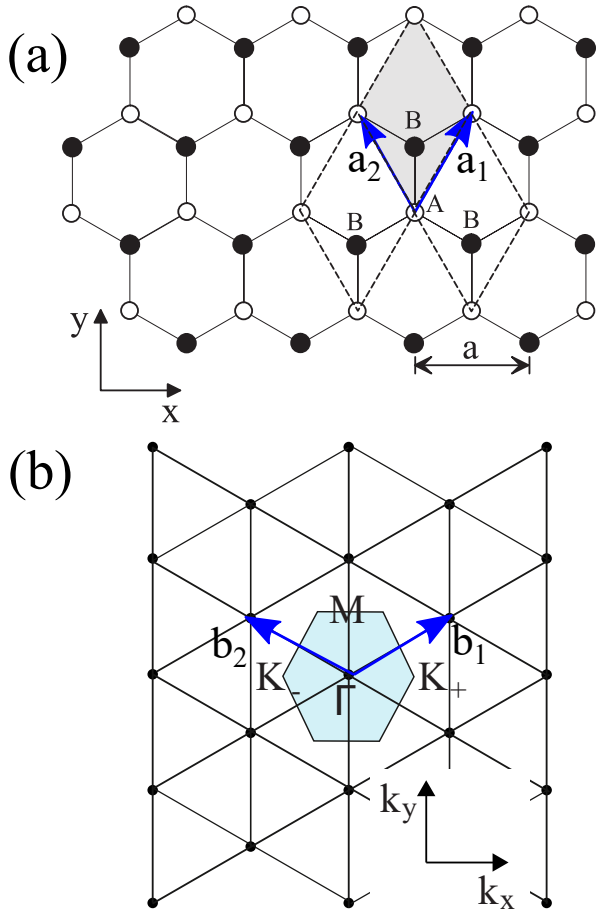


FIG. 2: (a) The honeycomb crystal structure of graphene consists of a hexagonal Bravais lattice with a basis of two atoms, labelled A and B, per unit cell (shaded rhombus). Unit cells of two other nearest-neighbour B atoms are shown as dotted rhombuses. (b) The reciprocal lattice (lattice points shown as dots) is a hexagonal Bravais lattice, the first Brillouin zone is a hexagon (shaded).

We assume that the dominant contribution arises when both orbitals are on the same site and use parameter ϵ_0 for this:

$$\begin{aligned} \epsilon_0 &= h_{AA}(\mathbf{R}=0) = \langle \phi(\mathbf{r}) \,|\mathcal{H}|\, \phi(\mathbf{r})\rangle, \\ H_{AA}(\mathbf{k}) &\approx h_{AA}(\mathbf{R}=0) = \epsilon_0. \end{aligned}$$

Since the B sublattice is identical to the A sublattice, and all the carbon atoms are chemically identical, $H_{BB} = H_{AA} \approx \epsilon_0$.

An off-diagonal transfer matrix element (9) is, e.g., given by

$$\begin{aligned} H_{AB}(\mathbf{k}) &= \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} h_{AB}(\mathbf{R}), \\ h_{AB}(\mathbf{R}) &= \langle \phi(\mathbf{r}) \,|\mathcal{H}|\, \phi(\mathbf{r}-\mathbf{R}-\mathbf{s}_B)\rangle. \end{aligned}$$

If we consider a given A atom, Fig. 2(a), it has three nearest-neighbour B atoms. They are located in unit

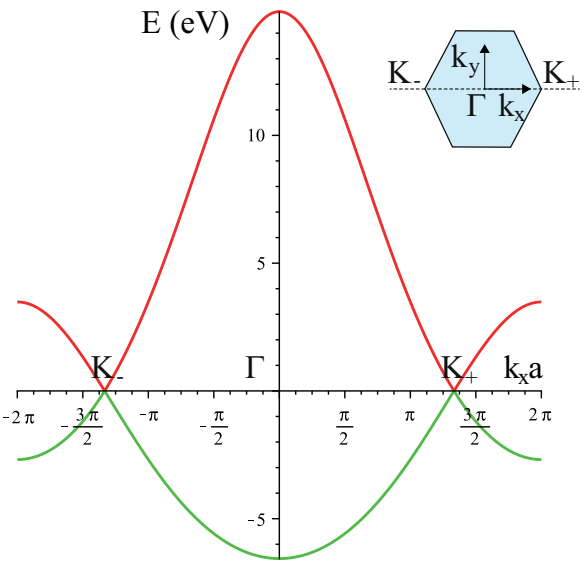


FIG. 3: The bands of graphene arising from the tight-binding model with one p_z orbital per atom, Eq. (19). The plot is made along a line in reciprocal space (dotted line in the inset) that intersects the centre, Γ, and two corners, K_+ and K_-, of the Brillouin zone. Parameter values are $\gamma_0 = 3.033\,\mathrm{eV}$, $\epsilon_0 = 0$ and $s_0 = 0.129$ [8].

cells given by $\mathbf{R} = 0$, $\mathbf{R} = -\mathbf{a}_1$, and $\mathbf{R} = -\mathbf{a}_2$. We describe these nearest-neighbour interactions with the same parameter γ_0:

$$-\gamma_0 = h_{AB}(0) = h_{AB}(-\mathbf{a}_1) = h_{AB}(-\mathbf{a}_2),$$

so that

$$\begin{aligned} H_{AB}(\mathbf{k}) &= -\gamma_0 \left(1 + e^{-i\mathbf{k}\cdot\mathbf{a}_1} + e^{-i\mathbf{k}\cdot\mathbf{a}_2}\right), \\ &= -\gamma_0 f(\mathbf{k}), \end{aligned}$$

with

$$f(\mathbf{k}) = 1 + 2e^{-ik_y\sqrt{3}a/2}\cos(k_x a/2)\,. \tag{16}$$

5. *Calculation of the energy*

To summarise, the transfer H and overlap S integral matrices of graphene are

$$H = \begin{pmatrix} \epsilon_0 & -\gamma_0 f(\mathbf{k}) \\ -\gamma_0 f^*(\mathbf{k}) & \epsilon_0 \end{pmatrix}, \tag{17}$$

$$S = \begin{pmatrix} 1 & s_0 f(\mathbf{k}) \\ s_0 f^*(\mathbf{k}) & 1 \end{pmatrix}. \tag{18}$$

(OK - I've briefly reinstated the overlap integral matrix S). It is possible to determine the energy by solving the secular equation (7):

$$E_\pm = \frac{\epsilon_0 \pm \gamma_0 |f(\mathbf{k})|}{1 \mp s_0 |f(\mathbf{k})|}\,. \tag{19}$$

The parameter values are listed by Saito *et al* [8] as $\gamma_0 = 3.033\,\mathrm{eV}$, $\epsilon_0 = 0$ and $s_0 = 0.129$.

The energy bands Eq. (19) are plotted in Fig. 3. At the corners of the Brillouin zone, labelled K_+ and K_- in Fig. 3, there is no band gap; the conduction and valence bands touch. This is because $f(\mathbf{k}) = 0$ at these points; there is no hopping between the A and B sublattices. As the A and B sublattices are identical, this causes a degeneracy at these points. The corners of the Brillouin zone are often referred to as 'valleys'. Although the Brillouin zone has six corners, only two of them are inequivalent, i.e., they can not be connected by a reciprocal lattice vector. Two corners K_+ and K_- with wave vectors $\mathbf{K}_\pm = \pm(4\pi/3a, 0)$ are labelled in Fig. 3. These crossing points are important because they coincide with the position of the Fermi level in pristine graphene. Moreover, the transfer matrix H has the form of a Dirac-like Hamiltonian near these points, describing massless chiral quasiparticles with a linear dispersion relation.

6. *Expansion near the K points*

We would like to derive a low-energy Hamiltonian in the vicinity of the K points. To do this, we introduce a momentum $\mathbf{p} = \hbar\mathbf{k} - \hbar\mathbf{K}_\xi$ measured with respect to the K point. Here the 'valley' index $\xi = \pm 1$ labels the two points K_+ and K_- with wave vectors $\mathbf{K}_\pm = \pm(4\pi/3a, 0)$, Fig. 3. We expand the function $f(\mathbf{k})$, Eq. (16), up to the first power of $\mathbf{p}$ which is valid for $pa/\hbar \ll 1$ where $p = |\mathbf{p}| = \sqrt{p_x^2 + p_y^2}$:

$$f(\mathbf{k}) \approx -\frac{\sqrt{3}a(\xi p_x - i p_y)}{2\hbar},$$

In this linear approximation, the overlap integral S may be replaced by a unit matrix, and the generalised eigenvalue equation (8) reduces to the Schrödinger equation $H\psi = E\psi$ with effective Hamiltonian

$$H = \begin{pmatrix} 0 & v(\xi p_x - i p_y) \\ v(\xi p_x + i p_y) & 0 \end{pmatrix}, \tag{20}$$

where $v = \sqrt{3}a\gamma_0/2\hbar$ is the band velocity. Here, we set $\epsilon_0 = 0$ to describe intrinsic graphene. The corresponding eigenenergy is $E = \pm v|\mathbf{p}|$, which approximately agrees with Eq. (19) for $pa/\hbar \ll 1$.

7. *Pseudospin and chirality*

Using Pauli spin matrices, the effective Hamiltonian Eq. (23) for $\xi = 1$ may be written as $H = v\boldsymbol{\sigma}.\mathbf{p}$ where $\boldsymbol{\sigma} = (\sigma_x, \sigma_y)$. The vector $\boldsymbol{\sigma}$ doesn't represent real electron spin, but a 'pseudospin' degree of freedom which describes the relative amplitude of the electron on the A and B sites of the honeycomb structure. If all the electron

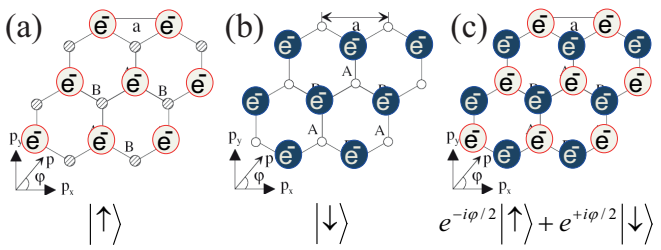


FIG. 4: Schematic illustration of pseudospin in graphene. (a) A pseudospin 'up' state $|\uparrow\rangle$ would have all electron density on the A sites and (b) a 'down' state $|\downarrow\rangle$ would have density on the B sites only. (c) In graphene, the pseudospin is a linear combination of up and down, and points in the plane of the graphene sheet.

density were situated on the A sites, this could be interpreted as a pseudospin 'up' state $|\uparrow\rangle = (1,0)$, Fig. 4(a), whereas a pseudospin 'down' state $|\downarrow\rangle = (0,1)$ would correspond to density solely on the B sites, Fig. 4(b). In graphene there are actually equal amounts of electron density on both the A and B sites, and the eigenstate of the effective Hamiltonian Eq. (23) is a linear combination of up and down pseudospin,

$$\psi = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm\xi e^{i\xi\varphi} \end{pmatrix} e^{i\mathbf{p}.\mathbf{r}/\hbar}, \tag{21}$$

$$= \frac{1}{\sqrt{2}} \left(|\uparrow\rangle \pm \xi e^{i\xi\varphi} |\downarrow\rangle \right) e^{i\mathbf{p}.\mathbf{r}/\hbar}, \tag{22}$$

where φ is the polar angle of momentum, $\mathbf{p} = (p_x, p_y) = p\,(\cos\varphi, \sin\varphi)$. Not only is there pseudospin, but the electrons are chiral: the direction of the pseudospin is fixed to the direction of the electronic momentum, either parallel or antiparallel.

The pseudospin in graphene, and its chiral nature (being fixed to the momentum direction), places a constraint on the scattering of electrons. A potential that doesn't break A-B symmetry is unable to influence the pseudospin which must, therefore, be conserved upon scattering. We can see this by evaluating the angular scattering probability which is proportional to $w(\varphi - \varphi') = |\langle\psi_\pm(\varphi)|\psi_\pm(\varphi')\rangle|^2 = \cos^2[(\varphi - \varphi')/2$, Fig. 5. It is anisotropic with an absence of backscattering $w(\pi) = 0$ [12, 13]. Scattering into a state with an opposite direction of momentum is prohibited because it would require a reversal of the direction of pseudospin, too. Pseudospin conservation is at the heart of many unusual transport properties in graphene including Klein tunneling which is anisotropic scattering at potential barriers [14, 15].

The Dirac-like effective Hamiltonian Eq. (20) is written for one valley $\xi = +1$ or -1 only; as such it is actually similar to a Weyl equation. In a four-component basis AK_+, BK_+, AK_-, BK_- it can be written as

$$H = v \begin{pmatrix} 0 & p_x - ip_y & 0 & 0 \\ p_x + ip_y & 0 & 0 & 0 \\ 0 & 0 & 0 & -p_x - ip_y \\ 0 & 0 & -p_x + ip_y & 0 \end{pmatrix}, \tag{23}$$

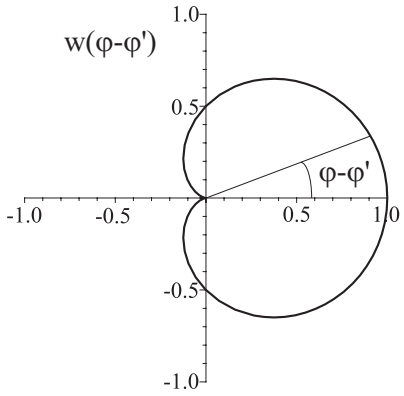


FIG. 5: Absence of backscattering in graphene, showing the angular dependence $w(\varphi - \varphi') = \cos^2[(\varphi - \varphi')/2]$ of the scattering probability off an A-B symmetric potential [12, 13].

which has the form of the Dirac equation. Just like the pseudospin related to the A/B sublattice degree of freedom, the valley degree of freedom $\xi = \pm 1$ can be viewed as a spin-like degree of freedom.

8. *Kane-Mele spin-orbit coupling in graphene*

We've seen that expansion in wave vector about a certain symmetry point in the Brillouin zone (e.g. Eq. (13) for the Γ point in a square lattice, Eqs. (20,23) for the K point in graphene) leads to a simple description with quadratic or linear dispersions. An alternative method to obtain this sort of description - rather than using the tight-binding model - would be to search for terms (in powers of wave vector) that satisfy the relevant global symmetries of the material in question.

This is what Kane and Mele did [16] in deducing the allowed spin-orbit terms in graphene. To follow this, we need to make things a bit more complicated by introducing (real) spin half. However, introducing spin threatens to turn the Dirac Hamiltonian (23) into an eight-component matrix (a space of 2 A/B lattices, 2 valleys, 2 spin components). Let's shorten notation by introducing 3 sets of 2×2 Pauli spin matrices:

$$\text{A/B lattice}: \quad \sigma_0, \sigma_x, \sigma_y, \sigma_z, \tag{24}$$

$$\text{valley}: \quad \Pi_0, \Pi_x, \Pi_y, \Pi_z, \tag{25}$$

$$\text{real spin}: \quad S_0, S_x, S_y, S_z, \tag{26}$$

where σ_0, Π_0, S_0 are 2×2 unit matrices in their respective spaces. The Dirac Hamiltonian (23) can now be written as

$$H = v\,(\Pi_z\sigma_x p_x + \Pi_0\sigma_y p_y)\,S_0. \tag{27}$$

We now consider symmetry operators. For example, the spatial inversion operator is $P = \Pi_x\sigma_x S_0$. Since momentum $\mathbf{p} \to -\mathbf{p}$ under spatial inversion, Hamiltonian (27) obeys $PH(p_x, p_y)P^{-1} = H(-p_x, -p_y)$. Simi-

larly, time reversal operation is given by $\Pi_x\sigma_0 S_y$ combined with complex conjugation. Again, momentum $\mathbf{p} \to -\mathbf{p}$ under time inversion, so Hamiltonian (23) obeys $TH(p_x, p_y)T^{-1} = H(-p_x, -p_y)$. Note that both operation P and T in graphene involve swapping the K points (hence you must take into account both K points when considering these symmetries).

Kane and Mele [16] were able to find a combination of matrices $\Pi_i\sigma_j S_\ell$ (with no momentum terms) that satisfied the symmetries of graphene including $P = \Pi_x\sigma_x S_0$ and $T = \Pi_x\sigma_0 S_y$ (plus complex conjugation). They found

$$H_{\rm KM} = \alpha\Pi_z\sigma_z S_z \tag{28}$$

where α is an unknown parameter [of course this kind of symmetry analysis (and the tight-binding model) is unable to say what the magnitude of the parameter α is]. It turns out that α is very small in graphene and, thus, this term is practically irrelevant. However, from a theoretical point of view, it is very important: it creates a special sort of band gap in graphene (more about this in the lectures on topological matter). Can you find any other invariants of the form $\Pi_i\sigma_j S_\ell$?

D. Summary

The tight-binding model provides a simple description (usually analytic) of electrons in various crystals, providing a Hamiltonian (i.e. the transfer integral matrix) and the band structure. If we want to calculate other properties of the material, e.g. transport properties, we could use the full tight-binding model as a starting point, although the subsequent calculations would usually have to be done numerically. However, it is often possible to expand for small wave vectors about a certain symmetry point to obtain a model of 'free' particles with an effective mass (e.g. Eq. (13) for the Γ point in a square lattice) or, sometimes, a linear dispersion (e.g. Eqs. (20,23) for the K point in graphene); whether this procedure is a reasonable approximation depends on many factors, not least of which is the position of the Fermi level, about which we've said nothing so far. In the following (when we discuss transport properties), we'll generally assume a quadratic dispersion with an effective mass.

III. MESOSCOPIC TRANSPORT IN THE DIFFUSIVE REGIME

A. Preliminaries

1. *Periodic boundary conditions and density of states*

As a simple example, consider a plane wave $\Psi(\mathbf{r}) = e^{i\mathbf{k}.\mathbf{r}}$ with a quadratic dispersion relation $E = \hbar^2k^2/2m$. Here $\mathbf{k} = (k_x, k_y)$ is the wave vector, $k = |\mathbf{k}| = \sqrt{k_x^2 + k_y^2}$.

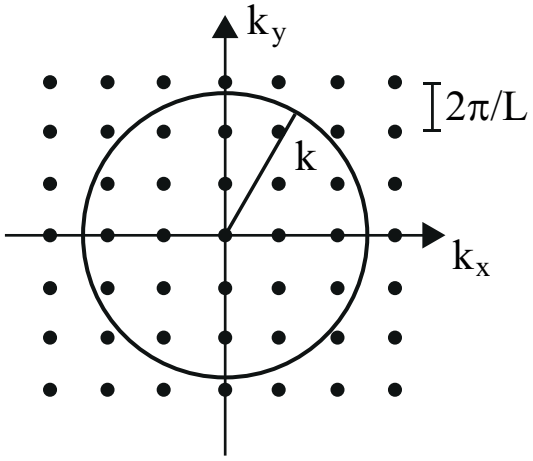


FIG. 6: Quantisation of wave vectors (black dots) for a box of length L with periodic boundary conditions, showing a circle of radius k in reciprocal space.

Sometimes it is useful to apply boundary conditions for a box of length L_x and L_y, in order to find quantisation conditions for the wave vector $\mathbf{k}$ and to normalise the wave function:

$$\begin{aligned} \Psi(x + L_x, y + L_y) &= \Psi(x, y), \\ e^{ik_x(x+L_x)+ik_y(y+L_y)} &= e^{ik_x x + ik_y y}, \\ \Rightarrow \qquad k_x = \frac{2\pi n_x}{L_x} \quad k_y &= \frac{2\pi n_y}{L_y}, \\ \Psi(x, y) &= \frac{1}{\sqrt{L_x L_y}} e^{ik_x x + ik_y y}. \end{aligned}$$

The quantisation of wave vector $\mathbf{k}$ helps us to introduce the concept of density of states. For a box of length L (i.e., $L_x = L_y = L$), the k values are separated by $2\pi/L$ and, thus, each state occupies an area $(2\pi/L)^2$ in reciprocal space, Fig. 6. The total area in reciprocal space for a circle of radius k is πk^2. Then, the number of states $N(k)$ within the circle of radius k is given by ratio of the total area to the area per state:

$$N(k) = \frac{\pi k^2}{(2\pi/L)^2} = \frac{k^2 L^2}{4\pi}. \tag{29}$$

The density of states $\nu(E)$ per unit area (L^2) per unit energy is

$$\begin{aligned} \nu(E) = \frac{g_s}{L^2}\frac{dN}{dE} &= \frac{g_s}{L^2}\frac{1}{(dE/dk)}\frac{dN}{dk} \\ &= \frac{g_s}{L^2}\left(\frac{m}{\hbar^2 k}\right)\frac{2kL^2}{4\pi} = \frac{g_s m}{2\pi\hbar^2}, \end{aligned} \tag{30}$$

where we included an extra degeneracy factor g_s ($g_s = 2$ for spin).

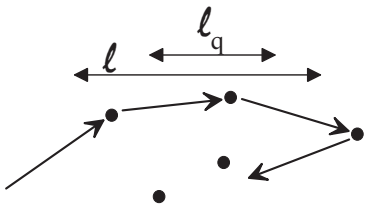


FIG. 7: Schematic drawing of the quantum scattering length ℓ_q and elastic mean free path ℓ in a disordered material.

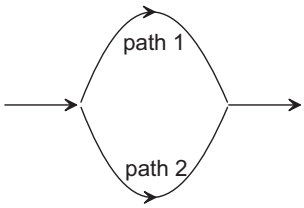


FIG. 8: Interference can be observed by splitting a beam of electrons and later recombining them.

2. Fermi function and electron density

At equilibrium, a state at energy E is occupied with probability given by the Fermi-Dirac distribution function $f(E)$ (or, simply, the Fermi function):

$$f(E) = \frac{1}{e^{(E-\mu)/k_B T} + 1}. \tag{31}$$

Here, the chemical potential μ (also called the Fermi level) is the energy for which the density of occupied states with higher energies is equal to the density of empty states with lower energies (technically its definition is the change in free energy when you add one electron to N electrons: $\mu = F_{N+1} - F_N$). The Fermi energy E_F is the energy of the highest occupied state at temperature $T = 0$, it coincides with the chemical potential μ at $T = 0$. The equilibrium electron density $n(E)$ (per unit area) takes into account the occupancy of the states:

$$n = \int_0^\infty \nu(E) f(E) dE. \tag{32}$$

3. Degenerate conductors

In these notes, we'll focus on the low temperature or 'degenerate' limit where $\mu \approx E_F$. Then,

$$f(E) \approx \Theta(\mu - E) = \begin{cases} 1 & E \le \mu, \\ 0 & E > \mu. \end{cases} \tag{33}$$

and the electron density is given by

$$n = \int_0^\infty \nu(E) f(E) dE = \int_0^\mu \nu(E) dE = \frac{g_s m \mu}{2\pi\hbar^2}. \tag{34}$$

Conduction properties are determined by electrons close to the Fermi energy, and their wave number is called the Fermi wave number, $E_F = \hbar^2 k_F^2/2m$ and $k_F = \sqrt{2mE_F/\hbar^2}$. Writing $n = g_s m E_F/(2\pi\hbar^2)$ shows that the Fermi wave number varies with the square root of the density: $k_F = \sqrt{4\pi n/g_s}$. The Fermi velocity v_F and Fermi momentum p_F are given by $v_F = \hbar^{-1}(dE/dk) = \hbar k_F/m = p_F/m$.

B. Characteristic scales

In this section, we'll describe some of the characteristic length and time scales in mesoscopic physics [2, 6].

1. Fermi wavelength

The Fermi wavelength λ_F is the de Broglie wavelength for charge carriers at low temperature. In two dimensions $\lambda_F = 2\pi/k_F = \sqrt{\pi g_s/n}$. For density $n = 5 \times 10^{11}\,\mathrm{cm}^{-2}$, then $\lambda_F \approx 35\,\mathrm{nm}$.

2. Momentum relaxation

For elastic scattering, such as electron-impurity scattering, the quantum scattering time τ_q is the average time between successive elastic scattering events, and the corresponding quantum scattering length $\ell_q = v_F \tau_q$ is the average distance between successive elastic scattering events, Fig. 7. Sometimes an electron only scatters by a small angle in a scattering event and, then, we use the concept of the momentum relaxation time which is longer than the quantum scattering time. The momentum relaxation time (or "Drude scattering time") τ is the average time before the initial momentum distribution is randomised, and the corresponding elastic mean free path $\ell = v_F \tau$ is the average distance travelled before the initial momentum distribution is randomised (i.e., the average distance between two strong scattering events). For density $n = 5 \times 10^{11}\,\mathrm{cm}^{-2}$ and $\tau = 100\,\mathrm{ps}$, then $v_F \approx 3 \times 10^5\,\mathrm{ms}^{-1}$ and $\ell \approx 30\,\mu\mathrm{m}$. In the semiclassical regime $\ell \gg \lambda_F$, it's useful to think of electron propagation in terms of Feynman's formulation of quantum mechanics, in which the amplitude for propagation is given by a sum over classical paths, each path having its own amplitude and phase.

3. The Aharonov-Bohm effect

To develop the concept of phase relaxation, we'll consider splitting a beam of electrons and later recombining them, Fig. 8. Imagine that, in a perfect crystal, the two

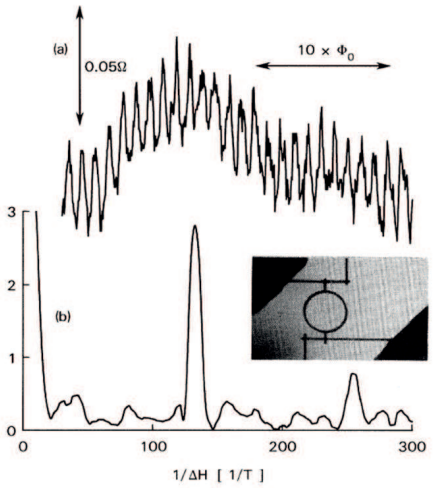


FIG. 9: Experimental data showing Aharonov-Bohm oscillations with period h/e in a gold ring at low temperature. Taken from Webb et al (1985) [18].

paths are identical. This results in constructive interference. A magnetic field could be applied to change the relative phase of the two paths, and alternate between constructive and destructive interference. This is the Aharonov-Bohm effect [17]: the phase of the wave function is influenced by the enclosed field, even though the field may be negligible along the path of the electron. The additional phase θ acquired by the electron when it moves in the region of vector potential $\mathbf{A}$ is given by the line integral along the path C:

$$\theta = -\frac{e}{\hbar}\int_C \mathbf{A}.\mathbf{d}\ell.$$

For a closed path, this phase may be related to the enclosed magnetic flux Φ using Stokes' Theorem:

$$\begin{aligned}\theta = -\frac{e}{\hbar}\oint_C \mathbf{A}.\mathbf{dl} &= -\frac{e}{\hbar}\int_S (\nabla\times\mathbf{A})\,.\mathbf{dS},\\ &= -\frac{e}{\hbar}\int_S \mathbf{B}.\mathbf{dS} = -\frac{2\pi\Phi}{\Phi_0},\end{aligned}$$

where the one-electron flux period is $\Phi_0 = h/e$. The transmission probability T for two electrons with relative amplitudes $e^{\pm i\theta/2}$ is given by

$$T = T_0\left|e^{i\theta/2} + e^{-i\theta/2}\right|^2 = T_0\left(1+\cos\theta\right), \qquad (35)$$

which displays oscillations as a function of magnetic field with period Φ_0. Such Aharonov-Bohm oscillations were observed in the mid-80s in the magnetoresistance of a gold ring at low temperature (0.01 K) [18], see Fig. 9.

The Aharonov-Bohm effect is significant in theoretical physics. In classical physics a magnetic field can only affect the behaviour of a charged particle if the particle moves through the field and experiences a force. The magnetic vector potential is only of significance inasmuch as it allows a convenient description of a force. By contrast, in quantum mechanics, the magnetic vector potential and the associated phase behaviour of the wave function appear to have a real, measurable effect. Or, to describe the effects of an electromagnetic field in terms of forces, one has to abandon the principle of locality.

4. Phase relaxation

Now, imagine performing this kind of interference experiment in a real crystal [2]. What will be the effect of different kinds of collisions? For example, what if the crystal contains impurities and defects? Then, the two paths will not be identical, so the interference may not be constructive for zero applied field. However, if the impurities and defects are static, there is a definite phase difference between the two paths (whether the interference is constructive or whatever). As the magnetic field is increased, there are alternative cycles of constructive and destructive interference. Thus, we conclude that static disorder doesn't cause phase relaxation.

Now, what if the crystal contains dynamic scatterers, such as phonons? In this case, the phase difference between the two paths varies randomly with time so that the interference fluctuates between constructive and destructive, time-averaging to zero. Interference can be observed only for electrons that are not scattered. Thus, the interference amplitude decays like $\exp(-t/\tau_\phi)$ where t is the transit time and τ_ϕ is the phase relaxation time. Thus, lattice vibrations (phonons) do cause phase relaxation.

What if we take into account electron-electron interactions? Electron-electron scattering causes dephasing because other electrons are not stationary. Note, however, that electron-electron scattering doesn't create momentum relaxation because the total momentum of the electron gas remains constant (momentum lost by one electron is gained by another).

Finally, what if the crystal contains magnetic impurities? Magnetic impurities have an internal spin degree of freedom that fluctuates with time, so they can cause phase-relaxation through spin-flip scattering. Note that their effect can be negated in a high magnetic field (bigger than k_BT) that suppresses spin flips and fluctuations.

The phase relaxation length L_ϕ is the distance travelled before the phase is randomised. In high mobility conductors, the motion of electrons in a phase-coherent region tends to be ballistic $\tau_\phi < \tau$, Fig. 10(a) and $L_\phi = v_F\tau_\phi$. In low mobility conductors, however, the motion of electrons in a phase-coherent region is diffusive $\tau_\phi > \tau$, Fig. 10(b). A typical phase-coherent diffusive path can be viewed as many (i.e. τ_ϕ/τ) ballistic segments each of length $\sim v_F\tau$. Then, the mean squared distance can be estimated as the sum of squares of lengths:

$$L_\phi^2 = \frac{\tau_\phi}{\tau}(v_F\tau)^2\langle\cos^2\theta\rangle = \frac{v_F^2\tau\tau_\phi}{2},$$

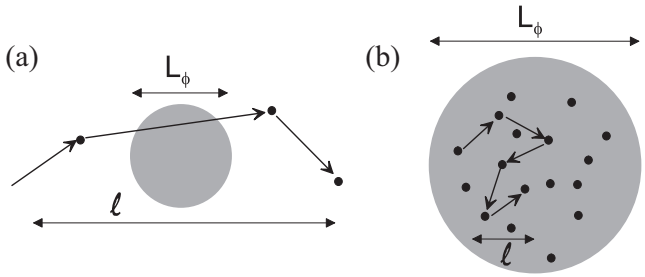


FIG. 10: Schematic drawing of electron transmission through a phase coherent region for (a) $L_\phi < \ell$ and (b) $L_\phi > \ell$.

$$L_\phi^2 \;=\; D\tau_\phi; \qquad\qquad D = \frac{v_F^2\tau}{2},$$

where D is the diffusion coefficient. Different regimes in mesoscopic conductors are summarised in Table I. In Section III D, we will describe interference effects in diffusive coherent samples $L_\phi \gtrsim L \gg \ell$. Before that, we will describe a classical equation of motion for the drift velocity in the diffusive regime.

TABLE I: Summary of different regimes in mesoscopic conductors as determined by the relation of the system size L to other characteristic length scales: ℓ elastic mean free path, L_ϕ phase relaxation length, and λ_F Fermi wavelength.

$L \gg \ell$ diffusive	$L \lesssim \ell$ ballistic
$L \gg L_\phi$ incoherent	$L \lesssim L_\phi$ phase coherent
$L \gg \lambda_F$ no size quantisation	$L \lesssim \lambda_F$ size quantisation

C. Classical regime

1. Classical equation of motion

To determine transport properties in the diffusive regime $L \gg \ell$, an appropriate tool would be the Boltzmann equation [6] in which the effect of electric field (say) on the Fermi distribution function (31) is taken into account perturbatively. As we have limited time here, we'll do something much easier, which is to consider a classical equation of motion for the drift velocity (essentially, this is the Drude model).

We can think of electrons as being classical particles that scatter in random directions from impurities, diffusive paths consist of a series of straight line (ballistic) segments [2, 6]. Generally, the speed along the straight line segments is given by the Fermi velocity $\mathbf{v}_F$ (which is the group velocity $\mathbf{v} = (1/\hbar)\nabla_{\mathbf{k}}E(\mathbf{k})$ at the Fermi surface). In the presence of an applied electric field $\mathbf{E}$ (note the clash of notation with energy E), electrons scatter in almost random directions, but with a drift velocity $\mathbf{v}_d$ in the direction of the net force: the drift velocity is related to current density $\mathbf{J}$ via a classical relation:

$$\mathbf{v}_d = -\frac{\mathbf{J}}{en}\,, \tag{36}$$

where the electric charge is $-e$ ($e > 0$).

The classical equation of motion for $\mathbf{v}_d$ is just Newton's Second Law with a friction term given by $-m\mathbf{v}_d/\tau$:

$$m\frac{d\mathbf{v}_d}{dt} = -e\left(\mathbf{E} + \mathbf{v}_d \times \mathbf{B}\right) - \frac{m\mathbf{v}_d}{\tau}. \tag{37}$$

For dc electric and magnetic fields $\mathbf{E}$, $\mathbf{B}$, there will be a transient solution and a steady state solution. The transient solution will just decay to zero over timescale τ and it's not of interest. We begin with $\mathbf{B} = 0$. The steady state solution can be found simply by setting $d\mathbf{v}_d/dt = 0$:

$$\mathbf{v}_d \;=\; -\frac{e\tau\mathbf{E}}{m}, \tag{38}$$

$$\mathbf{J} \;=\; -en\mathbf{v}_d = \frac{ne^2\tau\mathbf{E}}{m}, \tag{39}$$

$$\mathbf{J} \;=\; \sigma_0\mathbf{E} \quad\Rightarrow\quad \sigma_0 = \frac{ne^2\tau}{m}, \tag{40}$$

where σ_0 is the classical dc conductivity ('Drude' conductivity). Note that mobility μ is conductivity per unit charge density:

$$\mu = -\frac{\mathbf{v}_d}{\mathbf{E}} \equiv \frac{\sigma_0}{en} = \frac{e\tau}{m}\,. \tag{41}$$

Unfortunately, symbol μ is used for both mobility and chemical potential (mobility is usually given units of $cm^2/(Vs)$ 'centimetres squared per volt second').

2. Low field magnetoresistance

Now, in the equation of motion (37), we include a perpendicular magnetic field $\mathbf{B} = B\hat{\mathbf{k}}$, again set $d\mathbf{v}_d/dt = 0$ for steady state, and relate the drift velocity to the current density via $\mathbf{J} = -en\mathbf{v}_d$:

$$-\frac{mv_x}{e\tau} = E_x + v_yB \;\;; \qquad -\frac{mv_y}{e\tau} = E_y - v_xB,$$
$$-\frac{mv_x}{e\tau} - v_yB = E_x \;\;; \qquad -\frac{mv_y}{e\tau} + v_xB = E_y,$$
$$\frac{mJ_x}{e^2n\tau} + \frac{J_yB}{en} = E_x \;\;; \qquad \frac{mJ_y}{e^2n\tau} - \frac{J_xB}{en} = E_y.$$

These equations may be rewritten using $\sigma_0 = ne^2\tau/m$ and $\mu = \sigma_0/en$, and expressed as a matrix equation:

$$\frac{J_x}{\sigma_0} + \frac{\mu BJ_y}{\sigma_0} = E_x \;\;; \qquad \frac{J_y}{\sigma_0} - \frac{\mu BJ_x}{\sigma_0} = E_y,$$
$$\begin{pmatrix} E_x \\ E_y \end{pmatrix} \;=\; \frac{1}{\sigma_0}\begin{pmatrix} 1 & \mu B \\ -\mu B & 1 \end{pmatrix}\begin{pmatrix} J_x \\ J_y \end{pmatrix}.$$

Comparison with the resistivity tensor,

$$\begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{pmatrix} \rho_{xx} & \rho_{xy} \\ \rho_{yx} & \rho_{yy} \end{pmatrix}\begin{pmatrix} J_x \\ J_y \end{pmatrix},$$

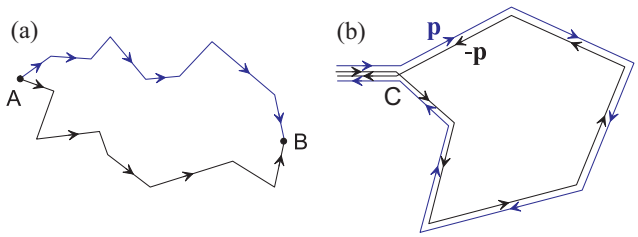


FIG. 11: Semiclassical paths contributing to interference effects in mesoscopic conductors (a) universal conductance fluctuations and (b) weak localisation.

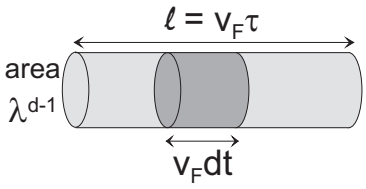


FIG. 12: For a ballistic segment of cross-sectional area $\sim \lambda_F^{d-1}$, the length travelled in time dt is $v_F dt$ and the corresponding volume is $v_F \lambda_F^{d-1} dt$.

shows that

$$\rho_{xx} = \rho_{yy} = \frac{1}{\sigma_0}, \tag{42}$$

$$\rho_{xy} = -\rho_{yx} = \frac{\mu B}{\sigma_0} = \frac{B}{en}. \tag{43}$$

The longitudinal resistivity $\rho_{xx} = \rho_{yy}$ is the inverse of the Drude conductivity, and the Hall resistivity is related to the inverse of the carrier density. Hall measurements are a vital experimental tool for determining the carrier density and the mobility.

3. *The Einstein model for degenerate conductors*

The Einstein model assumes that the net current is carried by electrons within a few $k_B T$ of the Fermi energy [2]. Diffusion is described by $\mathbf{J} = eD\nabla n$ leading to the Einstein relation for conductivity,

$$\sigma = e^2 \nu D. \tag{44}$$

This is consistent with the Drude conductivity (40) if

$$\begin{aligned} \frac{e^2 n \tau}{m} &= e^2 \nu D, \\ \frac{e^2 \nu m v_F^2 \tau}{2m} &= e^2 \nu D, \\ \Rightarrow \qquad D &= \frac{v_F^2 \tau}{2}, \end{aligned}$$

using $n = \nu E_F = \nu m v_F^2/2$, Eq. (34), for a degenerate conductor.

D. Interference effects in mesoscopic electronics

The equation of motion (37) treats electrons as classical particles, neglecting interference effects. In this section, we consider some of the interference effects.

1. *Weak localisation*

Weak localisation [19] is a correction to the classical dc conductivity arising from the quantum interference of electrons. To develop an understanding of it in the semiclassical regime $\ell \gg \lambda_F$, we can use Feynman's formulation of quantum mechanics, in which the amplitude for propagation is given by a sum over classical paths, each path having its own amplitude and phase $\varphi = \int \mathbf{k} \cdot \mathbf{dr}$. Typical classical paths connecting two points, A and B, are shown for a two-dimensional disordered conductor in Fig. 11(a). Each path follows the random walk of an electron experiencing diffusive motion; a series of straight-line segments connecting each scattering event. The scattering is elastic, so that the phase acquired along any given path is well defined. However, for propagation from A to B, there are many different paths in addition to the two shown. Each has a different length and acquired phase, so that the sum over them involves many random phases. Interference between such paths averages to zero and the mean electrical conductivity remains as described by the classical model.

Interference effects become relevant for closed paths because two electrons may start from a given point, C, and travel in opposite directions, Fig. 11(b). Two phases φ_1 and φ_2 acquired by the electrons as they propagate along the closed path in a clockwise and counter-clockwise direction are exactly equal. Thus, there is constructive interference of the counter-propagating electrons when they return to the starting point C. This enhances the probability to backscatter and produces a decrease of the conductivity. Traversal in two directions may be viewed as the motion of a particle and its time-reversed partner, indicating that their constructive interference is based on time reversal symmetry.

It is possible to determine the parametric dependence of the weak localisation correction using a 'back of an envelope' calculation. We will consider a diffusive path consisting of ballistic segments and estimate the probability that the path returns to its starting point with a certain time interval t to $t + dt$. To estimate the probability of return in time interval dt we consider the ratio of the volume explored in time interval t to $t + dt$ to the total volume explored by time t. We consider ballistic segments of the diffusive path to be of cross-sectional area $\sim \lambda_F^{d-1}$ for dimension d. Thus, the length travelled in time dt is $v_F dt$ and the corresponding volume is $v_F \lambda_F^{d-1} dt$ (see shaded region in Fig. 12).

For classical diffusion, the mean squared distance travelled after time t is given by $\langle r^2 \rangle = Dt$ so we can estimate

the total volume explored as $(Dt)^{d/2}$. Thus, the probability dP that the path returns to its starting point within a certain time interval t to $t+dt$ is

$$dP \approx \frac{v_F \lambda_F^{d-1} dt}{(Dt)^{d/2}}.$$

We see immediately that the probability depends on dimension d. To determine the overall probability we integrate with respect to time. Limits are set by the extent of the diffusive path, namely $t_{\min} = \tau$ and $t_{\max} = \min\{\tau_\phi, L^2/D\}$. In three dimensions,

$$\begin{aligned} P &\approx \frac{v_F \lambda_F^2}{D^{3/2}} \int_\tau^{\tau_\phi} \frac{dt}{t^{3/2}}, \\ &\approx \frac{v_F \lambda_F^2}{D^{3/2}} \left[\frac{1}{\sqrt{\tau}} - \frac{1}{\sqrt{\tau_\phi}} \right], \\ \frac{\delta g}{g} \approx -P &\approx -\frac{\lambda_F^2}{\ell^2} \left[1 - \sqrt{\frac{\tau}{\tau_\phi}} \right], \end{aligned}$$

where $g = G/(e^2/h)$ is the dimensionless conductance. The magnitude of the correction δg is small in the semi-classical regime $\ell \gg \lambda_F$ (hence 'weak' localisation) and, in $3d$, it is dominated by short trajectories and is finite even when $\tau_\phi \to \infty$.

In two dimensions,

$$\begin{aligned} P &\approx \frac{v_F \lambda_F}{D} \int_\tau^{\tau_\phi} \frac{dt}{t}, \\ &\approx \frac{v_F \lambda_F}{D} \ln\left(\frac{\tau_\phi}{\tau}\right) \sim \frac{v_F \lambda_F}{D} \ln\left(\frac{L_\phi}{\ell}\right), \\ \frac{\delta g}{g} \approx -P &\approx -\frac{v_F \lambda_F}{D} \ln\left(\frac{L_\phi}{\ell}\right). \end{aligned}$$

Using $g \approx h\nu D \sim D/(v_F \lambda_F)$ gives

$$\delta g \approx -\frac{2}{\pi} \ln\left(\frac{L_\phi}{\ell}\right), \tag{45}$$

where we have inserted the correct numerical factor. In contrast to $3d$, trajectories of all lengths contribute in $2d$ and the correction diverges for $\tau_\phi \to \infty$. The logarithm is an indication that the localisation behaviour of electrons in $2d$ (and lower dimensions) is different to that in higher dimensions. We will return to this in Section III.

Why is weak localisation useful to experimentalists? Like the Aharonov-Bohm effect, Section III B 3, weak localisation is affected by a perpendicular magnetic field. This leads to a characteristic low-field magnetoresistance which is observable. Consider two counter-propagating electrons as in Fig. 11(b) with amplitudes A_1 and A_2. The probability to return P is proportional to the modulus squared of the sum of the amplitudes:

$$P \sim |A_1 + A_2|^2 = |A_1|^2 + |A_2|^2 + A_1^* A_2 + A_1 A_2^*,$$

where the first two terms are the classical contributions of the individual paths; the second two terms describe interference. In the absence of a magnetic field, the phases and amplitudes are equal $A_2 = A_1 \equiv A$. In the presence of a perpendicular magnetic field, each path encloses a flux $\Phi = \int_S \mathbf{B}.\mathbf{dS}$ and, thus, they acquires phases as described by the amplitudes $A_1 = Ae^{-i\theta}$ and $A_2 = Ae^{i\theta}$ where $\theta = 2\pi\Phi/\Phi_0$ and $\Phi_0 = h/e$ is the one-electron flux period. Thus

$$P \sim |A_1 + A_2|^2 = 2|A|^2 + 2|A|^2 \cos 2\theta. \tag{46}$$

Remember - this is only the contribution of one given loop with two counter-propagating electrons. The net contribution to weak localisation is given by the sum of many different loops with different areas, different enclosed fluxes and, thus, different values of θ appearing as the argument of the cosine. Averaging over different loops destroys the interference term: a magnetic field destroys weak localisation, making it observable and useful. Suppressing weak localisation is equivalent to increasing the conductivity. In other words, a magnetic field breaks the time inversion symmetry in the system and leads to positive magnetoconductance: an increasing conductance with increasing field (alternatively, a negative magnetoresistance). Many early experiments are described in the review by Bergmann [20], typical data is shown in Fig. 13 [21]. The characteristic field scale for this suppression is the magnetic length $\ell_B = \sqrt{\hbar/eB}$ which is the typical spatial extent of a wave function in a magnetic field. At field $B = 1\,\mathrm{T}$, the magnetic length is already short, $\ell_B = 25.7\,\mathrm{nm}$, so the effect is generally observed at low fields $B < 1\,\mathrm{T}$.

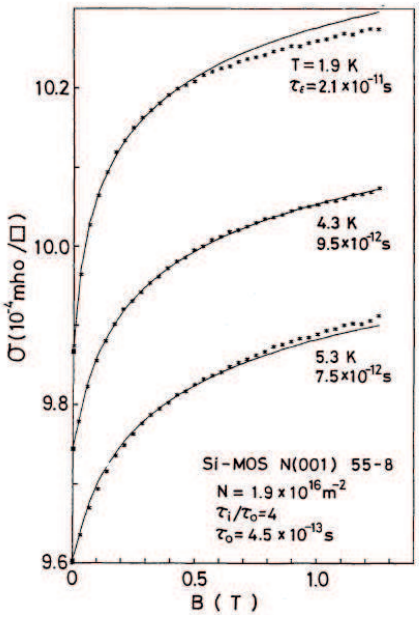


FIG. 13: Experimental data showing positive magnetoconductance due to weak localisation. Taken from Kawaji and Kawaguchi (1984) [21].

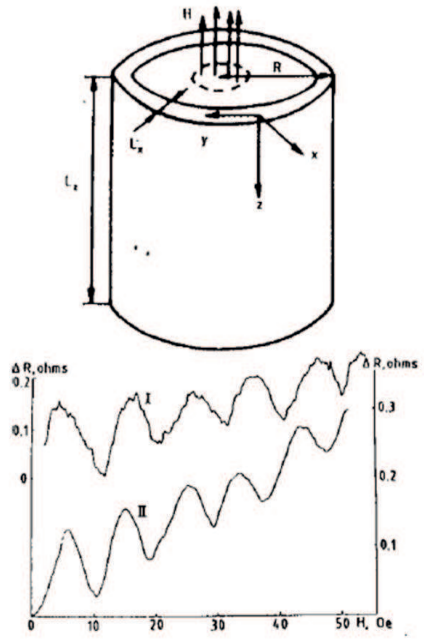


FIG. 14: Experimental data showing Altshuler-Aronov-Spivak oscillations with period $h/(2e)$. Taken from Sharvin and Sharvin (1981) [23].

2. *Altshuler-Aronov-Spivak (AAS) oscillations ($h/2e$)*

In weak localisation, both electrons traverse the whole loop, leading to a factor of 2θ in the sinusoidal term Eq. (46); the period is the superconducting flux period $h/(2e)$, half of the Aharonov-Bohm period Eq. (35). Magnetoresistance oscillations with this period were predicted by Altshuler et al [22] and observed by Sharvin and Sharvin [23] in hollow Mg cylinders of diameter $\sim 1.5\,\mathrm{mm}$, see Fig. 14. In this case, there were many different loops, but they were all constrained to the circumference of the hollow cylinder and, thus, they enclosed the same flux.

3. *Universal conductance fluctuations*

As mentioned above, Section III D 1, typical paths connecting two points, Fig. 11(a), don't contribute any correction to the average classical Drude conductivity. However, what about fluctuations of the conductivity? Consider two such paths with different amplitudes $A_1 = |A_1|e^{i\varphi_1}$, $A_2 = |A_2|e^{i\varphi_2}$, so their combined probability may be written as

$$\begin{aligned} P &= |A_1 + A_2|^2 = |A_1|^2 + |A_2|^2 + A_1^* A_2 + A_1 A_2^*, \\ &= |A_1|^2 + |A_2|^2 + 2|A_1||A_2|\cos(\varphi_1 - \varphi_2). \end{aligned}$$

The probability averaged over many paths is just given by the classical contribution $\langle P\rangle = |A_1|^2 + |A_2|^2$ because

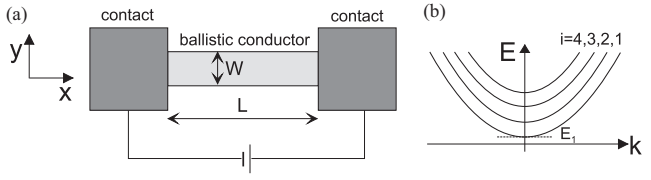


FIG. 15: (a) Schematic drawing of a $2d$ ballistic conductor of width W in the y direction, length L in the x direction, connected to bulk contacts. (b) Schematic of energies in the presence of quantisation in the y direction.

$\langle\cos(\varphi_1 - \varphi_2)\rangle = 0$. However, if we consider the square of the probability:

$$\begin{aligned} \langle P^2\rangle &= \left(|A_1|^2 + |A_2|^2\right)^2 + 4|A_1|^2|A_2|^2\langle\cos^2(\varphi_1 - \varphi_2)\rangle, \\ &= \left(|A_1|^2 + |A_2|^2\right)^2 + 2|A_1|^2|A_2|^2 \neq \langle P\rangle^2, \end{aligned}$$

since $\langle\cos^2(\varphi_1 - \varphi_2)\rangle = 1/2$. Thus, for the conductance $\langle G^2\rangle \neq \langle G\rangle^2$.

For *mesoscopic fluctuations* the properties of macroscopically identical samples that have different microscopic realisations of disorder are different, i.e., generally, for two samples $G_1 \neq G_2$. Every individual sample has a unique *magnetofingerprint* because its magnetoresistance [or magnetoconductance $G(B)$] depends on the specific microscopic realisation of disorder in it. We think in terms of an *ensemble average*: quantities are averaged with respect to many samples each with a different disorder realisation. Specifically, for the conductance $\langle\delta G^2\rangle = \langle(G - \langle G\rangle)^2\rangle = \langle G^2\rangle - \langle G\rangle^2 \sim (e^2/h)^2$. Or, in terms of the dimensionless conductance $g = G/(e^2/h)$ this is $\langle\delta g^2\rangle = \langle(g - \langle g\rangle)^2\rangle = \langle g^2\rangle - \langle g\rangle^2 \sim 1$. These are *universal conductance fluctuations (UCFs)* [24, 25]: here, universal means that the variance is independent of details specific to a particular sample including its conductance, but it does depend on dimensionality and global symmetries, such as time-inversion symmetry.

IV. THE LANDAUER-BÜTTIKER FORMULA

In this section, we describe the origin of the Landauer-Büttiker formula [26, 27] and explain how to use it in the context of scattering theory. On the face of it, the underlying concept is straightforward: the current through a conductor is related to the probability that an electron can travel through it.

A. Conductance quantisation

Ohm's law for the conductance of a sample of width W and length L is $G = \sigma W/L$. What happens as the length L gets shorter? Does the conductance become infinite for a ballistic conductor? If not, what is the source of the resistance?

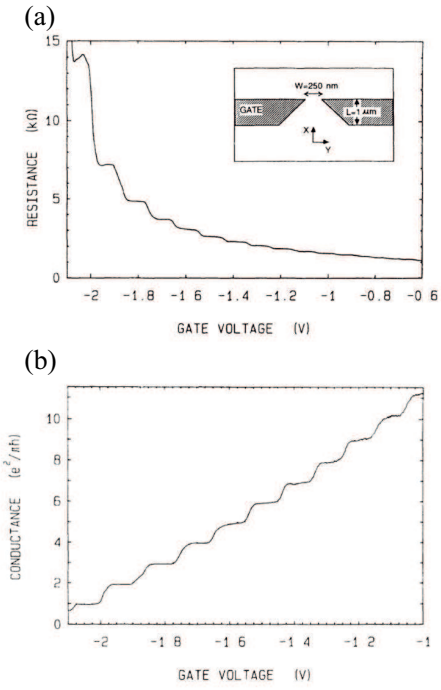


FIG. 16: Experimental data showing quantised conductance in a 2DEG on a GaAs-AlGaAs heterostructure. (a) Point-contact resistance as a function of gate voltage at 0.6 K (b) Conductance as a function of gate voltage, showing plateaus at multiples of $2e^2/h$. Taken from van Wees et al (1988) [28].

To begin to answer these questions, we consider a $2d$ ballistic conductor in the x-y plane [2] of width W in the y direction, length L in the x direction, connected to bulk contacts with a small bias voltage applied, Fig. 15(a). Owing to the boundaries in the y direction (e.g. with an infinite square-well confining potential), there are quantised energies corresponding to the y motion $E \approx E_i + \hbar^2 k^2/2m$, $i = 1, 2, \ldots$, Fig. 15(b) ($E_i = (\pi i\hbar/W)^2/(2m)$ for an infinite square well). When different subbands are created due to confinement in the x-y plane, they are often called 'transverse modes'. Using a split gate [see inset of Fig. 16(a)], it is possible to define a quantum point contact of variable width, and to observe what happens to the conductance as the width is varied. For clean enough samples, the result is quantised conductance, as observed by van Wees et al (1988) and Wharam et al (1988) [28, 29], Fig. 16. As the width is increased, additional transverse modes start to contribute to the conductance, and each of them contributes $2e^2/h$.

B. The Landauer formula

To introduce the Landauer-Büttiker formula, we consider transport through a narrow conductor [2]: a central conductor or 'scattering region' connected by two ballistic leads to large contacts across which the external bias is applied, Fig. 17(a). We assume that the large contacts

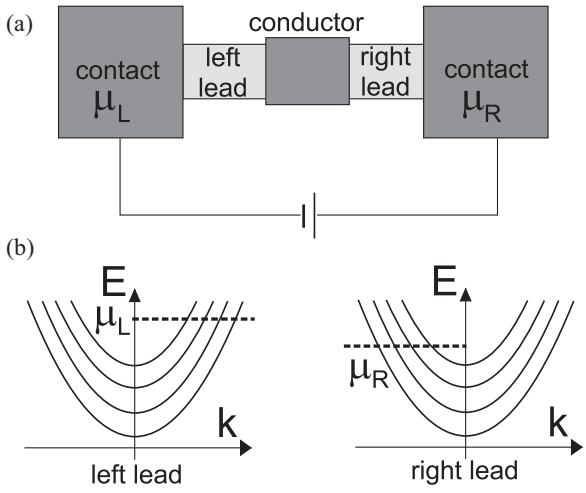


FIG. 17: (a) A central conductor is connected by two ballistic leads to large contacts across which the external bias is applied; $V = (\mu_L - \mu_R)/(-e)$ is the voltage drop across the conductor. (b) For reflectionless contacts, $+k$ states in the left lead must have the same 'quasi' Fermi level μ_L as the left contact and $-k$ states in the right lead have the same 'quasi' Fermi level μ_R as the right contact.

are much better conductors than the central conductor so that the voltage drop $V = (\mu_L - \mu_R)/(-e)$ is entirely across the central conductor. We also assume *reflectionless contacts*: the large contacts contain infinitely-many transverse modes, but there are only a few modes in the conductor and leads, thus an electron can exit from the lead into the contact with negligible probability of reflection (note, however, that an electron going from the contact to the lead has a large chance of reflection).

For reflectionless contacts, states with wave number $+k$ in the left lead must have originated from the left contact, and states with wave number $-k$ in the right lead must have originated from the right. Therefore, $+k$ states in the left lead must have the same 'quasi' Fermi level μ_L as the left contact; $-k$ states in the right lead have the same 'quasi' Fermi level μ_R as the right contact. At zero temperature, all states in the contacts are filled up to either μ_L or μ_R. Thus, current flow is due to electrons with energies in the range between μ_L and μ_R, and they must originate from the left contact.

To calculate the current, we consider electrons originating from the left contact in the energy range $\mu_L - \mu_R$. In the left lead, the current of electrons moving towards the conductor (with $+k$) is denoted $I_L^{\rm in}$, the current of electrons moving away from the conductor (with $-k$) is $I_L^{\rm out}$, and, in the right lead, the current moving away from the conductor (with $+k$) is $I_R^{\rm out}$, Fig. 18.

We consider a single transverse mode with $+k$ in the left lead and determine its contribution to the incoming current from the left, $I_L^{\rm in}$. For a plane wave in a $1d$ box of length L, normalisation $\int_0^L |\psi(x)|^2 dx = 1$ gives $\psi(x) = e^{ikx}/\sqrt{L}$, and the probability density per unit

length is $|\psi(x)|^2 = 1/L$ (obviously). The corresponding probability current is

$$j = -\frac{i\hbar}{2m}\left(\psi^* \frac{d\psi}{dx} - \psi \frac{d\psi^*}{dx}\right) = \frac{\hbar k}{mL} = \frac{v}{L}, \qquad (47)$$

where the velocity is $v = \hbar^{-1}(dE/dk) = \hbar k/m$. Then, to determine $I_L^{\rm in}$ we sum over all k values and remember the electronic charge $-e$:

$$I_L^{\rm in} = -\frac{e}{L}\sum_k v = -\frac{e}{L}\sum_k \frac{1}{\hbar}\frac{dE}{dk}.$$

It is convenient to convert the sum over k values into an integral. For a given function $f(k)$ with k values separated by Δ_k, then $\sum_k f(k)\Delta_k = \int f(k)dk$ in the limit $\Delta_k \to 0$. Since the separation in k is $2\pi/L$, then $\sum_k(\ldots) = (L/2\pi)\int(\ldots)dk$. Thus:

$$\begin{aligned} I_L^{\rm in} &= -\frac{e}{L}\sum_k \frac{1}{\hbar}\frac{dE}{dk} = -\frac{e}{\pi}\int \frac{1}{\hbar}\frac{dE}{dk}dk, \\ &= -\frac{2e}{h}\int_{\mu_R}^{\mu_L} dE = -\frac{2e}{h}\left(\mu_L - \mu_R\right), \end{aligned}$$

where a factor of 2 is included to account for spin degeneracy. Expressions for $I_L^{\rm out}$ and $I_R^{\rm out}$ are similar, but use the transmission probability T that an electron injected from the left lead will travel through the conductor into the right lead:

$$\begin{aligned} I_R^{\rm out} &= -\frac{2e}{h}T\left(\mu_L - \mu_R\right), \\ I_L^{\rm out} &= -\frac{2e}{h}\left(1-T\right)\left(\mu_L - \mu_R\right). \end{aligned}$$

Using these expressions, the total current is

$$\begin{aligned} I = I_L^{\rm in} - I_L^{\rm out} = I_R^{\rm out} &= -\frac{2e}{h}T\left(\mu_L - \mu_R\right), \\ &= \frac{2e^2}{h}T\frac{(\mu_L - \mu_R)}{-e}. \end{aligned}$$

Since $V = (\mu_L - \mu_R)/(-e)$ is the voltage drop across the conductor, the conductance $G = I/V$ takes a very simple form:

$$G = \frac{2e^2}{h}T. \qquad (48)$$

This is the Landauer Formula for a single-mode wire at zero temperature [26]; the conductance is a product of the 'quantum' of conductance $(2e^2/h)$ multiplied by the transmission probability T. The quantum of conductance $2e^2/h$ is equal to $7.75 \times 10^{-5}\,\Omega^{-1}$. The von Klitzing constant is a quantum of resistance $R_K = h/e^2 = 25813\,\Omega$, named after the discoverer of the integer quantum Hall effect.

Consider a ballistic conductor, with ideal transmission $T = 1$, then $G = 2e^2/h$. Where does the resistance come from [2, 3]? We can rearrange the Landauer formula with an arbitrary value of T,

$$\begin{aligned} G^{-1} &= \frac{h}{2e^2}\frac{1}{T} = \frac{h}{2e^2}\left(\frac{1-T}{T} + \frac{T}{T}\right), \\ &= \frac{h}{2e^2}\frac{R}{T} + \frac{h}{2e^2} = G_s^{-1} + G_c^{-1}, \end{aligned}$$

so that the resistance is written as a sum of resistances in series. Here $R = 1 - T$ is the reflection probability. The resistance G_c^{-1} (where $G_c = 2e^2/h$) may be interpreted as a contact resistance. We assumed that the contacts are reflectionless for electrons passing from lead to contact. However, for the opposite direction, for electrons to pass from the contact to the lead, there is a large probability of reflection, and, thus, a non-zero contact resistance. The resistance G_s^{-1} (where $G_s = (2e^2/h)T/R$) may be interpreted as the resistance of the central conductor. As expected, $G_s \to \infty$ as $T \to 1$. Note that the conductance $G_s = (2e^2/h)T/R$ is the original form of the conductance formula derived by Landauer [26].

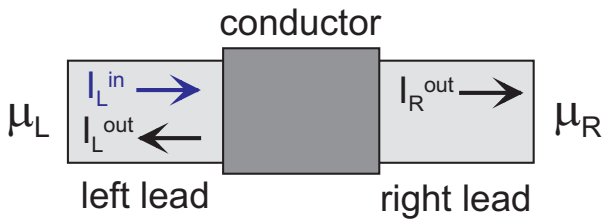


FIG. 18: In the left lead, the current of electrons moving towards the conductor (with $+k$) is denoted $I_L^{\rm in}$, the current of electrons moving away from the conductor (with $-k$) is $I_L^{\rm out}$, and, in the right lead, the current moving away from the conductor (with $+k$) is $I_R^{\rm out}$.

C. Scattering theory in one dimension

In order to use the Landauer formula, it is necessary to calculate the transmission probability T. This may be done using scattering theory [2, 6, 30]. In general, the scattering region (the central conductor) may be described by a complicated Hamiltonian (e.g., it may include disorder), but we assume that the leads are ballistic and described by the Schrödinger equation with constant potentials V_L, V_R, respectively. The wave functions $\psi_L(x)$ and $\psi_R(x)$ in the leads are

$$\psi_L = Ae^{ik_L x} + Be^{-ik_L x};\ k_L = \sqrt{\frac{2m(E-V_L)}{\hbar^2}}, \qquad (49)$$

$$\psi_R = Ce^{ik_R x} + De^{-ik_R x};\ k_R = \sqrt{\frac{2m(E-V_R)}{\hbar^2}}. \qquad (50)$$

We would like to determine the relationships between the amplitudes A, B, C, D. It is sufficient to consider what

happens to unit amplitude incoming from the left, i.e. set $A = 1$, $B = r$, $C = t$, $D = 0$, where r and t are reflection and transmission coefficients:

$$\psi_L^{\rm in} = e^{ik_L x}; \qquad \psi_L^{\rm out} = re^{-ik_L x}; \qquad \psi_R^{\rm out} = te^{ik_R x}.$$

Then, the transmission and reflection probabilities are defined as T = transmitted flux/incoming flux and R = reflected flux/incoming flux. We define flux using the probability current, Eq. (47), so that a state of the form $\psi = Ne^{ikx}$ has a corresponding flux equal to $j = v|N|^2$. Since $v = \hbar k_{L/R}/m$,

$$T = \frac{k_R}{k_L}|t|^2; \qquad R = |r|^2. \tag{51}$$

1. Transfer matrix

A transfer matrix M relates states in the right lead to those in the left lead:

$$\begin{pmatrix} C \\ D \end{pmatrix} = M \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix}. \tag{52}$$

In our case,

$$\begin{pmatrix} t \\ 0 \end{pmatrix} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \begin{pmatrix} 1 \\ r \end{pmatrix}, \tag{53}$$

$$\Rightarrow \quad t = \frac{\det(M)}{M_{22}} = \frac{k_L}{k_R M_{22}}; \quad T = \frac{k_L}{k_R |M_{22}|^2}, \tag{54}$$

where $\det(M) = k_L/k_R$.

2. Transfer matrix - example

As a simple example, we consider tunnelling through a constant potential barrier [6, 30] of height $V > E$ located at $0 < x < L$, Fig. 19(a). We have solutions to the Schrödinger equation in three regions, left lead ψ_L, right lead ψ_R, and the central scattering region ψ_s:

$$\begin{aligned} \psi_L &= Ae^{ikx} + Be^{-ikx}; \qquad k = \sqrt{\frac{2mE}{\hbar^2}}, \\ \psi_R &= Ce^{ikx} + De^{-ikx}, \\ \psi_s &= Fe^{\eta x} + Ge^{-\eta x}; \qquad \eta = \sqrt{\frac{2m(V-E)}{\hbar^2}}. \end{aligned}$$

To solve this problem, we use boundary condutions at the interfaces ($x = 0$ and $x = L$) which are the usual matching conditions for the Schrödinger wave functions, namely 1) continuity of the wave function ψ and 2) continuity of the first derivative of ψ. It is convenient to calculate the transfer matrices M_1 and M_2 at each interface ($x = 0$ and $x = L$) separately, and then to combine them:

$$\begin{pmatrix} F \\ G \end{pmatrix} = M_1 \begin{pmatrix} A \\ B \end{pmatrix}; \qquad \begin{pmatrix} C \\ D \end{pmatrix} = M_2 \begin{pmatrix} F \\ G \end{pmatrix},$$

$$\Rightarrow \qquad M = M_2 M_1; \qquad \begin{pmatrix} C \\ D \end{pmatrix} = M \begin{pmatrix} A \\ B \end{pmatrix}.$$

Note that this procedure of combining transfer matrices may be generalised to many sections $M = M_N \ldots M_3 M_2 M_1$ (which is why transfer matrices are useful). We find

$$\begin{aligned} M_1 &= \frac{1}{2\eta} \begin{pmatrix} ik+\eta & -ik+\eta \\ -ik+\eta & ik+\eta \end{pmatrix}, \\ M_2 &= \frac{1}{2ik} \begin{pmatrix} (ik+\eta)e^{-(ik-\eta)L} & (ik-\eta)e^{-(ik+\eta)L} \\ (ik-\eta)e^{(ik+\eta)L} & (ik+\eta)e^{(ik-\eta)L} \end{pmatrix}. \end{aligned}$$

Using $M = M_2 M_1$ and $T = 1/|M_{22}|^2$ gives

$$\begin{aligned} T &= \left[1 + \left(\frac{k^2+\eta^2}{2k\eta}\right)^2 \sinh^2 \eta L\right]^{-1}, \\ &= \left[1 + \frac{V^2}{4E(V-E)} \sinh^2 \eta L\right]^{-1}, \end{aligned}$$

as plotted in Fig. 19(b). For analytic calculations, transfer matrices are very useful and intuitive since $M = M_N \ldots M_3 M_2 M_1$. They are used in many areas of physics, e.g. optics. However, for numerical calculations, scattering matrices are generally more stable.

3. Scattering matrix

A scattering matrix $\widetilde{S}$ relates outgoing states to incoming states:

$$\begin{pmatrix} B \\ C \end{pmatrix} = \widetilde{S} \begin{pmatrix} A \\ D \end{pmatrix} = \begin{pmatrix} \widetilde{S}_{11} & \widetilde{S}_{12} \\ \widetilde{S}_{21} & \widetilde{S}_{22} \end{pmatrix} \begin{pmatrix} A \\ D \end{pmatrix}. \tag{55}$$

In our case,

$$\begin{aligned} \begin{pmatrix} r \\ t \end{pmatrix} &= \begin{pmatrix} \widetilde{S}_{11} & \widetilde{S}_{12} \\ \widetilde{S}_{21} & \widetilde{S}_{22} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \\ \Rightarrow \quad t &= \widetilde{S}_{21}; \quad r = \widetilde{S}_{11}, \\ T &= \frac{k_R}{k_L}|\widetilde{S}_{21}|^2; \quad R = |\widetilde{S}_{11}|^2. \end{aligned}$$

This $\widetilde{S}$ matrix is written in terms of wave amplitudes. However, we'd like to write it in terms of flux (i.e. current) amplitudes. So, we define:

$$\begin{aligned} S_{ij} &= \sqrt{\frac{v_i}{v_j}}\widetilde{S}_{ij}, \\ \begin{pmatrix} \sqrt{v_L}B \\ \sqrt{v_R}C \end{pmatrix} &= S \begin{pmatrix} \sqrt{v_L}A \\ \sqrt{v_R}D \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} \sqrt{v_L}A \\ \sqrt{v_R}D \end{pmatrix}. \end{aligned}$$

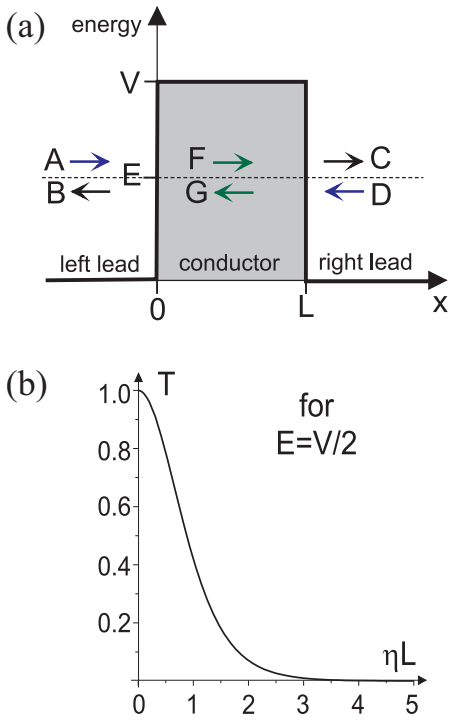


FIG. 19: (a) a simple example of a scattering region: a constant potential barrier of height V located at $0 < x < L$. (b) The corresponding transmission probability for energy $E = V/2$, plotted as a function of barrier length L.

Now, there is a neater relationship between the transmission and reflection coefficients and the S matrix: $T = |S_{21}|^2$ and $R = |S_{11}|^2$.

4. *Scattering matrix and current conservation*

Current conservation places a constraint on the form of the scattering matrix. We begin by writing the form of the Hermitian adjoint of S:

$$\begin{pmatrix} \sqrt{v_L}B \\ \sqrt{v_R}C \end{pmatrix} = S \begin{pmatrix} \sqrt{v_L}A \\ \sqrt{v_R}D \end{pmatrix}, \tag{56}$$

$$\begin{pmatrix} \sqrt{v_L}B^* & \sqrt{v_L}C^* \end{pmatrix} = \begin{pmatrix} \sqrt{v_L}A^* & \sqrt{v_L}D^* \end{pmatrix} S^\dagger. \tag{57}$$

For states in the left and right leads, Eqs. (49,50), the probability current, Eq. (47), is given by

$$j_L = v_L|A|^2 - v_L|B|^2; \qquad j_R = v_R|C|^2 - v_R|D|^2.$$

Current conservation $j_R = j_L$ gives

$$v_L|B|^2 + v_R|C|^2 = v_L|A|^2 + v_R|D|^2. \tag{58}$$

We can express the left hand side of this as

$$\text{l.h.s.} = v_L|B|^2 + v_R|C|^2,$$

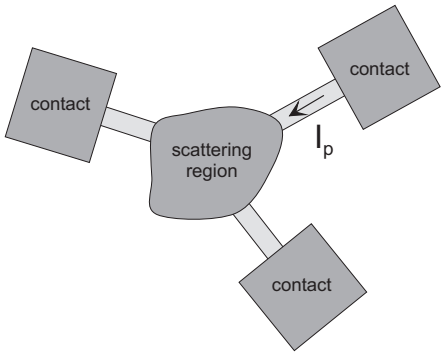


FIG. 20: The Landauer-Büttiker formula describes a system with multi-mode leads and many contacts.

$$= \begin{pmatrix} \sqrt{v_L}B^* & \sqrt{v_R}C^* \end{pmatrix} \begin{pmatrix} \sqrt{v_L}B \\ \sqrt{v_R}C \end{pmatrix},$$

$$= \begin{pmatrix} \sqrt{v_L}A^* & \sqrt{v_R}D^* \end{pmatrix} S^\dagger S \begin{pmatrix} \sqrt{v_L}A \\ \sqrt{v_R}D \end{pmatrix},$$

where we used Eqs. (56,57). Thus, the left hand side of Eq. (58) is equal to the right hand side $v_L|A|^2 + v_R|D|^2$ if $S^\dagger S = 1$. In other words, current conservation requires the S matrix to be unitary. $S^\dagger S = 1$ implies that $|S_{11}|^2 + |S_{21}|^2 = 1$, i.e., $R + T = 1$ where $T = |S_{21}|^2$ and $R = |S_{11}|^2$.

D. The Landauer-Büttiker formula

The Landauer-Büttiker formula [26, 27] is a generalisation to include multi-mode leads and many contacts, Fig. 20. It is possible to generalise the Landauer formula by expressing the current in a given lead p in terms of a sum over all other leads:

$$I_p = -\frac{2e}{h} \sum_q T_{pq} \left(\mu_p - \mu_q\right),$$

where T_{pq} is the transmission function for electron transfer from lead q to lead p:

$$T_{pq} = \sum_{m,n} |t^{pq}_{mn}|^2.$$

Many modes per lead are taken into account by t^{pq}_{mn} which is the flux amplitude for an electron to leave via mode m in lead p if there were unit incoming flux amplitude in channel n of lead q.

V. LOCALISATION THEORY

In Section I, we described the semiclassical regime $\ell \gg \lambda_F$ and discussed quantum corrections, such as weak

localisation, to the classical dc conductivity. They are generally well understood and of relevance to mesoscopic conductors. In this section, we'll consider the situation of strong disorder; what happens when $\ell \sim \lambda_F$? Semiclassics breaks down - is it possible for the electronic wave function to be localised by a potential that varies randomly in space (strong localisation)? We'll describe the scaling theory of localisation, which provides a connection between the regimes of weak and strong localisation. Throughout, we'll assume that electrons are non-interacting and phase coherent over the whole sample.

A. Anderson localisation

1. *Anderson tight-binding model of localisation*

Ideas about strong localisation were developed by Anderson [31] in the 1950s, and it is often called 'Anderson localisation'. In his Nobel Prize lecture of 1977, Anderson said "Localization ..., very few believed it at the time, and even fewer saw its importance, among those who failed to fully understand it at first was certainly its author. It has yet to receive adequate mathematical treatment, and one has to resort to the indignity of numerical simulations to settle even the simplest questions about it."

To begin, let's consider a one-electron tight-binding Hamiltonian, Fig. 21(a), with one orbital $|i\rangle$ per lattice site $i = 1, 2, \ldots, N$ of energy ε_i, and with hopping integral t connecting nearest neighbours,

$$H = \sum_{i=1}^{N} \varepsilon_i |i\rangle\langle i| - t \sum_{\langle i,j \rangle}^{N} \left(|i\rangle\langle j| + |j\rangle\langle i| \right). \qquad (59)$$

In a perfect crystal, all the on-site energies ε_i are equal and the eigenstates may be written as a linear combination of atomic orbitals,

$$|\Phi_k\rangle = \frac{1}{\sqrt{N}} \sum_i e^{i\mathbf{k}.\mathbf{r}_i} |i\rangle, \qquad (60)$$

where $\mathbf{k}$ is a good quantum number. The band width is proportional to the hopping integral t. Explicitly, in one dimension:

$$\begin{aligned} H &= \varepsilon \sum_i |i\rangle\langle i| - t \sum_i \left(|i\rangle\langle i+1| + |i+1\rangle\langle i| \right), \\ \langle m|H|n\rangle &= \varepsilon \delta_{m,n} - t \left(\delta_{m+1,n} + \delta_{m,n+1} \right), \\ |\Phi_k\rangle &= \frac{1}{\sqrt{N}} \sum_n e^{ikna} |n\rangle, \\ E(k) &= \langle \Phi_k | H | \Phi_k \rangle = \varepsilon - 2t \cos(ka), \end{aligned}$$

where a is the lattice constant. The tight-binding model of the two-dimensional square lattice was discussed in Section II B.

To consider localisation [3, 4], we take into account site-diagonal disorder. A common model is that on-site energies ε_i are uncorrelated random variables drawn from a rectangular probability distribution over the range $-W/2 < \varepsilon_i < W/2$, Fig. 21(b). Parameter W produces site-energy mismatch that acts as a barrier to propagation/delocalisation, whereas t favours propagation/delocalisation. The dimensionless ratio W/t is a measure of the degree of disorder. To understand the role of the ratio W/t, consider a system with just two sites with different on-site energies $\varepsilon_2 - \varepsilon_1 = 2W$,

$$H = \begin{pmatrix} -W & -t \\ -t & W \end{pmatrix}; \qquad \Psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$$

The energies correspond to a bonding $E_- = -\sqrt{W^2 + t^2}$ and anti-bonding $E_+ = \sqrt{W^2 + t^2}$ pair. In the limit $W/t \to 0$, the states are delocalised with weight on both sites, $\Psi_\pm = (1, \mp 1)/\sqrt{2}$, whereas for $W/t \to \infty$ they are localised $\Psi_- = (1, 0)$, $\Psi_+ = (0, 1)$.

In three dimensions, there is a critical value $(W/t)_c$ of the disorder parameter, such that, for $W/t > (W/t)_c$, all eigenstates are exponentially localised and the system is an insulator - the 'Anderson insulator'. Exponential localisation means that the envelope of an eigenstate decays exponentially with distance far away from the position $\mathbf{r}_\alpha$ at which it peaks:

$$|\psi_\alpha\rangle = \sum_i a_i^\alpha |i\rangle; \qquad |a_i^\alpha| \sim \exp\left(-\frac{|\mathbf{r}_i - \mathbf{r}_\alpha|}{\xi_\alpha} \right).$$

Fig. 22(a) is a schematic of a localised state with amplitude envelope decaying exponentially over the localisation length ξ. The wave function peaks near a site that it is in near-resonance with, $E_\alpha \approx \varepsilon_i$. As a contrast, Fig. 22(b) is a schematic of an extended state in a disordered system.

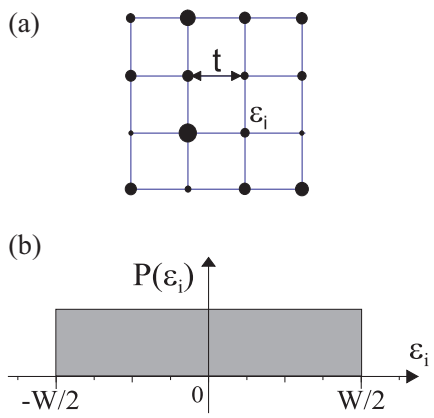


FIG. 21: (a) The Anderson tight-binding model of localisation for a square lattice with nearest-neighbour hopping integral t and random on-site energies ε_i (b) On-site energies ε_i are uncorrelated random variables drawn from a rectangular probability distribution over the range $-W/2 < \varepsilon_i < W/2$.

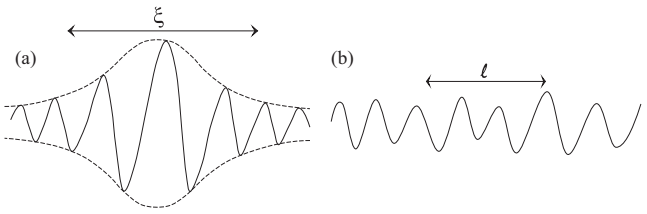


FIG. 22: (a) Schematic of a localised state with amplitude envelope decaying exponentially over the localisation length ξ and (b) schematic of an extended state.

2. Fluctuations

Theoretically at least, we often think of an ensemble of samples [3, 4]: macroscopically identical, but microscopically different, e.g., all samples have the same density of impurities, but the distribution of impurities is different from sample-to-sample. *Ensemble average*: averaging with respect to an ensemble made up of a very large number of samples. Ensemble averaging is relatively easy to do theoretically, but generally very hard to do experimentally. *Parametric variations*: unfortunately, experiments are usually only performed on one or two samples so ensemble averaging is impossible. However, it is often possible to take measurements on one sample while varying a parameter e.g. chemical potential, magnetic field, quantum dot shape; then perform an averaging with respect to the parametric variations. *Ergodicity*: ensemble average yields the same results as averaging with respect to variations of a suitably chosen parameter.

The local density of states,

$$\nu_i(E) = \sum_\alpha |\langle i|\alpha\rangle|^2 \delta(E - E_\alpha), \tag{61}$$

counts the eigenstates ψ_α of energy $E_\alpha = E$ having significant overlap with (i.e., localised near to) the ith site. The global density of states

$$\nu(E) = \sum_\alpha \delta(E - E_\alpha), \tag{62}$$

counts all the eigenstates of energy $E_\alpha = E$ irrespective of their position in space. *Self-averaging*: in any given random and large sample, the global density of states takes the same value as its ensemble average. The local density of states is not self-averaging: for localised systems, there are a small number of states that are localised near the ith site, and $\nu_i(E)$ will fluctuate from sample to sample, depending on the microscopic realisation of disorder.

3. Sub-critical disorder

If all states are localised in three dimensions for $W/t > (W/t)_c$, what happens for sub-critical disorder $W/t < (W/t)_c$? This was discussed by Mott [32]: states in the middle of the band are extended (not localised), whereas states at the band extremities are localised. The two sets of states are separated by a sharply defined energy called the mobility edge, Fig 23. *Metal-insulator transition*: the system undergoes a metal-insulator transition as the Fermi level crosses the mobility edge even though the density of states is continuous and non-zero across the mobility edge. Experiments on microwave cavities [33] have observed strongly localised and delocalised states at different frequencies in disordered billiards, Fig. 24. *Anderson transition*: as disorder increases, the mobility edges E_{c1} and E_{c2} close in and eventually meet at the critical disorder $(W/t)_c$ when all states are localised.

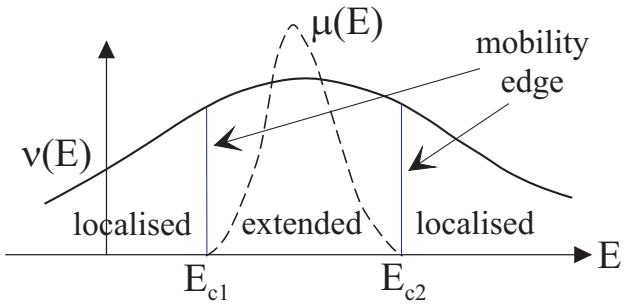


FIG. 23: For sub-critical disorder $W/t < (W/t)_c$, states in the middle of the band are extended, whereas states at the band extremities are localised. The two sets of states are separated by a sharply defined energy called the mobility edge.

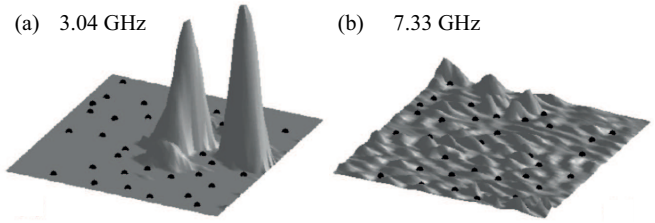


FIG. 24: Experimental data showing measured eigenstates of a disordered microwave cavity billiard with 36 scatterers (shown as black dots): (a) a strongly localized state, $f = 3.04\,\mathrm{GHz}$, and (b) a delocalized state $f = 7.33\,\mathrm{GHz}$. Taken from Pradhan and Sridhar (2000) [33].

B. Scaling theory

1. Thouless picture

A scaling theory of localisation was developed by Thouless [34]. Imagine we have a d-dimensional cube with side of length L at zero temperature. Is it a conductor or insulator? Or - a better question - as the system size is increased towards infinity, does it evolve into a conductor or insulator? How do initial conditions - at

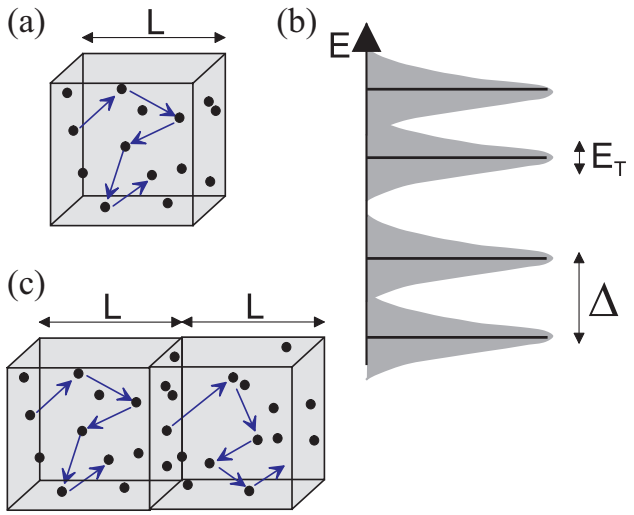


FIG. 25: Thouless picture. The ratio g of the Thouless energy E_T to the mean level spacing Δ determines the nature of states as d-dimensional cubes are joined together to make a larger system.

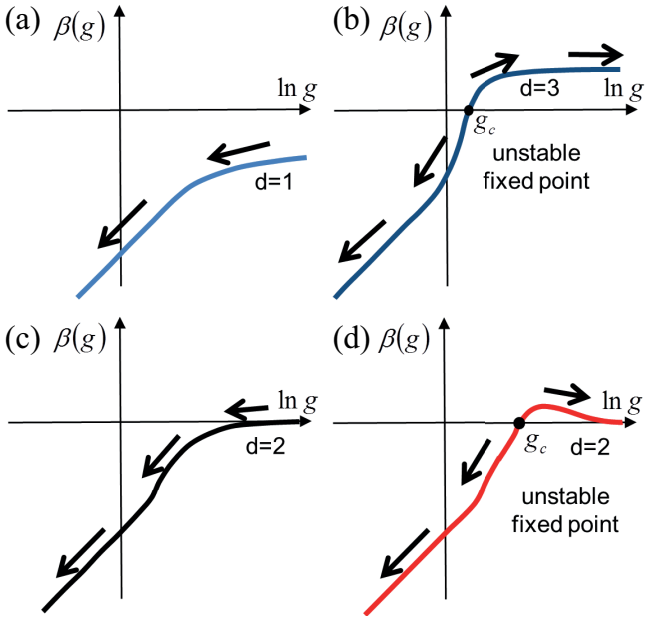


FIG. 26: The scaling function $\beta(g)$ in different dimensions (a) one dimension, (b) three dimensions, (c) two dimensions, (d) two dimensions in the presence of strong spin-orbit coupling. Black arrows show the direction of flow as the length L increases.

the initial length scale - determine whether it evolves into a conductor or insulator? That is, for a finite-size sample, if we know the size-dependence of its conductance, does this tell us about the conductor-versus-insulator tendency of the sample?

We consider a d-dimensional cube, Fig. 25(a). The time taken for an electron to diffuse across the cube is $\tau_T = L^2/D$ where D is the diffusion coefficient. The *Thouless energy* E_T is the corresponding energy,

$$E_T = \frac{h}{\tau_T} = \frac{hD}{L^2}. \tag{63}$$

This should be compared, Fig. 25(b), with the *mean level spacing* Δ for a finite cube, which is the inverse of the density of states,

$$\Delta = \frac{1}{L^d \nu(E)}. \tag{64}$$

The ratio g of the Thouless energy to the mean level spacing tells us what happens as the scale L increases:

$$g = \frac{E_T}{\Delta}. \tag{65}$$

What happens when we start to join cubes together, Fig. 25(c)? For $g < 1$, $E_T < \Delta$, there is little overlap between levels in adjacent cubes; the electron can't travel easily between cubes so the states tend to be localised. For $g > 1$, $E_T > \Delta$, levels from different cubes overlap; the electrons can travel between cubes easily so the states are extended.

The dimensionless parameter $g = E_T/\Delta$ determines the scaling behaviour. What is its physical meaning? Diffusion is related to conductivity through the Einstein relation $\sigma = e^2\nu D$, and we can use this to rewrite the Thouless energy,

$$\begin{aligned} E_T &= \frac{hD}{L^2} = \frac{h\sigma}{e^2\nu L^2} = \frac{h\sigma L^d \Delta}{e^2 L^2}, \\ \Rightarrow \quad g &= \frac{E_T}{\Delta} = \frac{h\sigma L^{d-2}}{e^2} = \frac{h}{e^2} G. \end{aligned}$$

Since $G = \sigma L^{d-2}$ is the conductance, $g = E_T/\Delta$ is the dimensionless conductance (i.e., conductance in units of e^2/h). *One parameter scaling*: the idea that $g = E_T/\Delta$ is the only relevant scaling variable.

2. *The gang of four*

The so-called gang of four, Abrahams, Anderson, Licciardello and Ramakrishnan (1979) [35], introduced the scaling ansatz: it is possible to interpolate between the limits of a good conductor and a localised insulator. The interpolation is cast in the form of a differential equation,

$$\beta(g) = \frac{d\ln g}{d\ln L}. \tag{66}$$

We assume *universality*: the scaling function $\beta(g)$ should be independent of details specific to a particular material. However, it may depend on dimensionality and global symmetries, e.g., time-inversion symmetry. Next, we deduce the form of the scaling function $\beta(g)$ in the limits of weak and strong disorder. For weak disorder, $g \gg 1$,

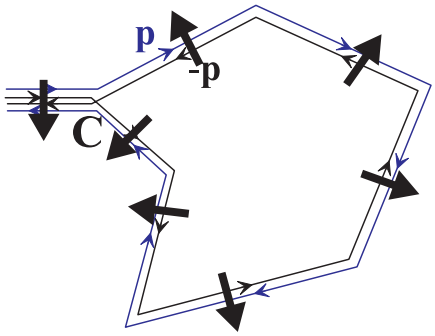


FIG. 27: Semiclassical paths contributing to weak localisation, showing spin orientation (black arrows) in the presence of strong spin-orbit coupling.

we have metallic conduction $g \sim \sigma L^{d-2}$:

$$\begin{aligned} g \sim L^{d-2} \quad &\Rightarrow \quad \ln g \approx (d-2)\ln L, \\ &\Rightarrow \quad \beta(g) = d-2. \end{aligned}$$

The sign of β for $g \gg 1$ depends on the dimension. For strong disorder, $g \ll 1$, $g \sim g_0 e^{-\alpha L}$:

$$\begin{aligned} g \sim g_0 e^{-\alpha L} \quad &\Rightarrow \quad \ln g \approx \ln g_0 - \alpha L, \\ &\Rightarrow \quad \beta(g) = \frac{d(-\alpha L)}{d\ln L}, \\ x = \ln L \quad &\Rightarrow \quad \beta(g) = \frac{d(-\alpha e^x)}{dx} = -\alpha e^x, \\ &\beta(g) = -\alpha L = \ln g - \ln g_0. \end{aligned}$$

The sign of β for $g \ll 1$ is always negative. Now we know the value of $\beta(g)$ in the limits of weak and strong disorder, we extrapolate to all values of g by assuming β is a well-behaved, smooth function.

Fig. 26(a) shows the form of $\beta(g)$ in one dimension. It is always negative. This means that g will decrease with increasing L; as it does so, $\beta(g)$ becomes even more negative. This continues until g is zero. No matter what the conductance is in a finite-size sample, the system will always tend to insulating behaviour as the scale is increased; the direction of flow is shown by the black arrows in Fig. 26(a). There is no metal-insulator transition in $1d$ because all states are localised.

In three dimensions, the scaling function is positive for $g \gg 1$, negative for $g \ll 1$, with an unstable fixed point at $\beta(g) = 0$, Fig. 26(b). There is a metal-insulator transition in $3d$. For $g > g_c$, the flow is to conducting behaviour $g \gg 1$ and $g \sim L$. For $g < g_c$, the flow is to insulating behaviour $g \ll 1$ and $g \sim g_0 e^{-\alpha L}$. The fixed point g_c is unstable, meaning the system flows away from it. If $g = g_c$ on some scale, then $g = g_c$ on all scales so that the conductivity $\sigma \sim g_c/L \to 0$.

Two dimensions is a marginal case because $\beta(g) = 0$ for $g \gg 1$: does it approach zero from above or below [as in Fig. 26(d) and (c)]? The answer is given by the weak localisation correction: $\beta(g) = -C/g$. where $C > 0$. Let's check this:

$$\begin{aligned} \frac{d\ln g}{d\ln L} &= -\frac{C}{g}, \\ \int g d\ln g &= -C\int d\ln L, \\ x = \ln g \quad \Rightarrow \quad \int e^x dx &= -C\int d\ln L, \\ g &= \text{const} - C\ln L. \end{aligned}$$

which is the weak localisation correction, Eq. (45). Thus, $\beta(g) = -C/g$ for $g \gg 1$, so the scaling function $\beta(g)$ approaches zero from below as in Fig. 26(c). In $2d$, the system is never a metal (there is no transition in $2d$) '2d metals are not really metals'. However, although the system is never a true metal, a given sample can be a good conductor in the weak localisation regime for $L < \xi$. For $L \approx \xi$, the system enters the strongly localised regime.

3. *Spin-orbit coupling*

Strong Rashba spin-orbit coupling can influence weak localisation in two dimensions, Fig 27. Rashba spin-orbit tends to couple spin to the direction of momentum so, as the electron traverses a closed path, the spin rotates. Spin rotation introduces an additional phase difference between the electrons as they propagate along the closed path in different direction, and strong Rashba spin-orbit coupling tends to drive weak localisation into weak anti-localisation, i.e. $g = \text{const} + C\ln L$. Experimental data from Bergmann (1984) [20] is shown in Fig. 28.

Spin-orbit coupling influences the scaling theory: weak anti-localisation means that the scaling function is positive for $g \gg 1$, $\beta(g) = +C/g$, so the scaling function $\beta(g)$ approaches zero from above as in Fig. 26(d). So, spin-orbit coupling leads to a metal-insulator transition in $2d$.

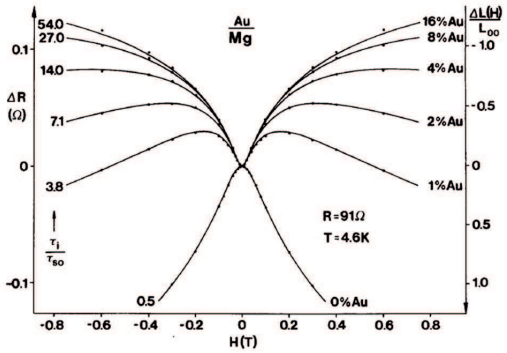


FIG. 28: Experimental data (points) of the low-field magnetoresistance of a thin Mg film with different coverages of Au. Solid lines show fits to theory by Hikami et al [36]. Taken from Bergmann (1984) [20].

VI. SUMMARY

These notes provide a very brief introduction to some concepts in mesoscopic physics. Section II was a description of the tight-binding model with a focus on graphene and Section III described mesoscopic transport in the diffusive regime including the classical equation of motion for the drift velocity and quantum interference effects especially weak localisation. Then, Section IV described the Landauer-Büttiker formula, scattering theory in $1d$, transfer and scattering matrices. Localisation theory was covered in Section V: Anderson, Thouless, and the Gang of Four.

There are many interesting and important topics not covered including high-field magnetoresistance, Shubnikov-de-Haas, quantum Hall effect; edge states, spin Hall effect, topological insulators; quantum chaos, Random Matrix Theory, semiclassics; quantum dots, Coulomb Blockade; mesoscopic superconductivity, Andreev reflection, proximity effect; spintronics; diagrammatic perturbation theory, field theory; optics; etc. We refer the interested reader to other texts, e.g., [1–6], for more details.

[1] C. W. J. Beenakker and H. van Houten, *Quantum Transport in Semiconductor Nanostructures* in *Solid State Physics, vol. 44* eds. H. Ehrenreich and D Turnbull (New York, Academic Press, 1991).

[2] S. Datta, *Electronic Transport in Mesoscopic Systems* (Cambridge University Press, Cambridge, 1995).

[3] Y. Imry, *Introduction to Mesoscopic Physics* (Oxford University Press, Oxford, 2nd Ed., 2001).

[4] P. A. Mello and N. Kumar, *Quantum Transport in Mesoscopic Systems* (Oxford University Press, Oxford, 2004).

[5] E. Akkermans and G. Montambaux, *Mesoscopic Physics of Electrons and Photons* (Cambridge University Press, Cambridge, 2007).

[6] T. Heinzel, *Mesoscopic Electronics in Solid State Nanostructures* (Wiley-VCH, Weinham, 3rd Ed., 2010).

[7] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Brooks/Cole, Belmont CA, 1976).

[8] R. Saito, M. S. Dresselhaus and G. Dresselhaus, *Physical Properties of Carbon Nanotubes* (London: Imperial College Press, 1998).

[9] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, Science **306**, 666 (2004).

[10] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, M. I. Katsnelson, I. V. Grigorieva, S. V. Dubonos, A. A. Firsov, Nature **438**, 197 (2005).

[11] Y. B. Zhang, Y. W. Tan, H. L. Stormer, P. Kim, Nature **438**, 201 (2005).

[12] T. Ando, T. Nakanishi, and R. Saito, *J. Phys. Soc. Jpn.* **67**, 2857 (1998).

[13] P. L. McEuen, M. Bockrath, D. H. Cobden, Y.-G. Yoon, and S. G. Louie, *Phys. Rev. Lett.* **83**, 5098 (1999).

[14] V. V. Cheianov and V. I. Falko, Phys. Rev. B **74**, 041403(R) (2006).

[15] M. I. Katsnelson, K. S. Novoselov, and A. K. Geim, Nature Phys. **2**, 620 (2006).

[16] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 146802 (2005).

[17] Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).

[18] R. A. Webb, S. Washburn, C. P. Umbach, and R. B. Laibowitz, Phys. Rev. Lett. **54**, 2696 (1985).

[19] B. L. Altshuler, D. Khmelnitski, A. I. Larkin, and P. A. Lee, Phys. Rev. B **22**, 5142 (1980).

[20] G. Bergmann, Phys. Rep. **107**, 1 (1984).

[21] S. Kawaji and Y. Kawaguchi JPSP **53**, 2868 (1984).

[22] B. L. Altshuler, A. G. Aronov, and B. Z. Spivak, JETP Lett. **33**, 94 (1981).

[23] D. Y. Sharvin and Y. V. Sharvin, JETP Lett. **34**, 272 (1981).

[24] B. L. Altshuler, JETP Lett. **41**, 648 (1985).

[25] P. A. Lee and A. D. Stone, Phys. Rev. Lett. **55**, 1622 (1985).

[26] R. Landauer, Phil. Mag. **21**, 863 (1970).

[27] M. Büttiker, IBM J. Res. Dev. **32**, 317 (1988).

[28] B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. van der Marel, and C. T. Foxon, Phys. Rev. Lett. **60**, 848 (1988).

[29] D. A. Wharam, T. J. Thornton, R. Newbury, M. Pepper, H. Ahmed, J. E. F. Frost, D. G. Hasko, D. C. Peacock, D. A. Ritchie, and G. A. C. Jones, J. Phys. C: Solid State Physics, **21**, L209 (1988).

[30] E. Merzbacher, *Quantum Mechanics* (Wiley, 3rd Ed., 1998).

[31] P. W. Anderson, Phys. Rev. **109**, 1492 (1958).

[32] N. F. Mott, Phil. Mag. **13**, 989 (1966).

[33] P. Pradhan and S. Sridhar, Phys. Rev. Lett. **85**, 2360 (2000).

[34] D. J. Thouless, Phys. Rep. **13**, 93 (1974).

[35] E. Abrahams, P. W. Anderson, D. C. Licciardello, and T. V. Ramakrishnan, Phys. Rev. Lett. **42**, 673 (1979).

[36] S. Hikami, A. I. Larkin, and Y. Nagaoka, Prog. Theor. Phys. **63**, 707 (1980).

Problems

1. **Tight-binding model of the Kagome lattice**

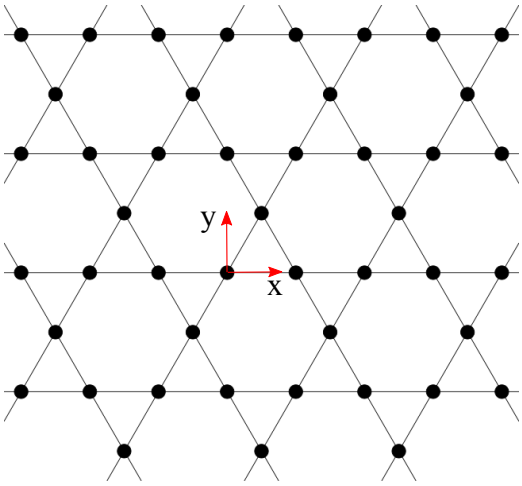


Figure 1: The two-dimensional Kagome lattice.

Figure 1 shows the Kagome lattice where all nearest-neighbour separation distances (shown as solid lines) are equal. Assume the atoms are chemically identical with one s-orbital each.

(a) Denoting the lattice constant as a and using the Cartesian coordinate system shown in the figure, give expressions for two primitive lattice vectors $\mathbf{a}_1$ and $\mathbf{a}_2$. Draw the lattice vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ on the lattice and identify a primitive unit cell on the lattice.

(b) Give expressions for the primitive reciprocal lattice vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ (corresponding to your choice of $\mathbf{a}_1$ and $\mathbf{a}_2$ in (a)). On a separate sketch, indicate several reciprocal lattice points and indicate the first Brillouin zone.

(c) Determine the tight-binding Hamiltonian for this system in $\mathbf{k}$ space taking nearest-neighbour hopping into account. You may assume that diagonal elements are zero, and that the tight-binding parameter for nearest-neighbour hopping is t.

(d) Determine the energy eigenvalues of your tight-binding Hamiltonian in terms of a, t, k_x and k_y, using the fact that there is a dispersionless band at energy $E = -2t$. [Hint: (i) use the following identity for arbitrary A, B:

$$\cos^2(A) + \cos^2(B) + \cos^2(A \pm B) = 1 + 2\cos(A)\cos(B)\cos(A \pm B)$$

(ii) You should have a cubic equation for energy E. Use the fact there is a solution $E = -2t$ to factor out $E + 2t$, yielding a quadratic equation for the other two E values.]

(e) Expand your expressions for energy near the Γ point ($k_x = k_y = 0$) to derive simple expressions for energy as a power series in k_x and k_y up to quadratic order. Hence, sketch all the energy bands (including $E = -2t$) in the vicinity of the Γ point. [Hint: the simple expressions should be isotropic in $\mathbf{k}$-space].

2. **Classical equation of motion for ac conductivity.**

Consider a two-dimensional disordered conductor in the x-y plane in the presence of a static perpendicular magnetic field $\mathbf{B} = (0, 0, B)$ and an in-plane ac electric field with frequency ω and components

$$E_x(t) = E_{x0}\, e^{i\omega t}\,; \qquad E_y(t) = E_{y0}\, e^{i\omega t}.$$

(a) Using $\mathbf{v}_d = \mathbf{v}_{d0}\, e^{i\omega t}$ for drift velocity and $\mathbf{J} = \mathbf{J}_0\, e^{i\omega t}$ for current density, use the classical equation of motion for $\mathbf{v}_d$ to show that the resistivity tensor is given by

$$\rho = \frac{1}{\sigma_0} \begin{pmatrix} 1 + i\omega\tau & \omega_c\tau \\ -\omega_c\tau & 1 + i\omega\tau \end{pmatrix},$$

where τ is the momentum scattering time and $\omega_c = eB/m$ is the cyclotron frequency.

(b) Hence determine the conductivity tensor.

(i) Simplify the conductivity tensor for zero frequency $\omega\tau = 0$. How does the magnetic field dependence of the components compare with that of the resistivity?

(ii) Simplify the conductivity tensor in the high frequency limit $\omega \gg \omega_c$ and $\omega\tau \gg 1$. Using the plasma frequency $\omega_p = \sqrt{ne^2/(\varepsilon m)}$, ε is a constant, show that

$$\sigma_{xx} = \sigma_{yy} = -i\frac{\varepsilon\omega_p^2}{\omega}\,; \qquad \sigma_{xy} = -\sigma_{yx} = \frac{\varepsilon\omega_c\omega_p^2}{\omega^2}\,.$$

3. **Dirac-like Hamiltonian of graphene.** The effective low-energy Hamiltonian of monolayer graphene in the vicinity of a single valley is given by

$$H_0 = v \begin{pmatrix} 0 & \xi p_x - i p_y \\ \xi p_x + i p_y & 0 \end{pmatrix}, \tag{1}$$

where $p_x = -i\hbar\partial_x$, $p_y = -i\hbar\partial_y$ are momentum operators, $\xi = \pm 1$ denotes the valley, the centre of which is at wave vector $\mathbf{K}_\xi = \xi\,(4\pi/3a, 0)$, a is the lattice constant. You may assume that the Hamiltonian acts on column vectors

$$\Psi = \begin{pmatrix} \psi_A \\ \psi_B \end{pmatrix} e^{i\mathbf{k}.\mathbf{r}}\,, \tag{2}$$

where the wave vector is $\mathbf{k} = (k_x, k_y)$ and ψ_A, ψ_B describe wave function amplitudes on A and B sites that are independent of position $\mathbf{r} = (x, y)$.

(a) Determine the energy eigenvalues E of the eigenvalue equation $H_0\Psi = E\Psi$ with the Dirac-like Hamiltonian Eq. (1).

(b) Hence, determine the energy dependence of the density of states (per unit area per unit energy) in graphene.

(c) Using polar coordinates for the wave vector $\mathbf{k} = k\,(\cos\phi, \sin\phi)$, determine the column vector Ψ, Eq. (2), of the eigenvalue equation $H_0\Psi = E\Psi$, and normalise it so that $|\psi_A|^2 + |\psi_B|^2 = 1$. Hence, determine the expectation values $\langle\sigma_x\rangle$, $\langle\sigma_y\rangle$, $\langle\sigma_z\rangle$ of the Pauli spin matrices with respect to Ψ. How is the direction of the vector $\langle\boldsymbol{\sigma}\rangle = \langle\sigma_x\rangle\hat{\boldsymbol{\imath}} + \langle\sigma_y\rangle\hat{\boldsymbol{\jmath}} + \langle\sigma_k\rangle\hat{\boldsymbol{k}}$ related to the direction of the wave vector?

4. **Landau levels.** In the presence of a perpendicular magnetic field, the energy spectrum of a 2DEG is quantised into Landau levels. The spectrum may be calculated using an appropriate gauge, e.g. the Landau gauge, whereby it is possible to reduce the problem to a $1d$ harmonic oscillator with wave functions corresponding to harmonic oscillator states ϕ_n. Such states may be expressed in terms of Hermite polynomials and exponential functions. As this is a standard text book exercise, we'll just use the result here. For convenience, we express the momentum operators as $\pi = p_x + ip_y$ and $\pi^\dagger = p_x - ip_y$. In the presence of a perpendicular magnetic field, they are proportional to lowering and raising operators of the harmonic oscillator states ϕ_n:

$$\pi\phi_n = -\frac{\sqrt{2}i\hbar}{\lambda_B}\sqrt{n}\,\phi_{n-1}\,, \tag{3}$$

$$\pi^\dagger\phi_n = \frac{\sqrt{2}i\hbar}{\lambda_B}\sqrt{n+1}\,\phi_{n+1}\,, \tag{4}$$

where $\pi\phi_0 = 0$ and the magnetic length is $\lambda_B = \sqrt{\hbar/eB}$. The cyclotron frequency is given by $\omega_c = eB/m$. Determine the Landau level spectrum for the following Hamiltonians, and compare the dependence of the energies on level index n and magnetic field B:

(a) Free electrons in two dimensions:

$$H = \frac{1}{4m}(\pi^\dagger\pi + \pi\pi^\dagger)\,.$$

(b) Graphene in the vicinity of one of the valleys:

$$H_1 = v\begin{pmatrix} 0 & \pi^\dagger \\ \pi & 0 \end{pmatrix}.$$

(c) Bilayer graphene in the vicinity of one of the valleys:

$$H_2 = -\frac{1}{2m}\begin{pmatrix} 0 & \left(\pi^\dagger\right)^2 \\ \pi^2 & 0 \end{pmatrix}.$$

5. **Transfer matrix for a delta function.** Consider a one-dimensional system with zero potential everywhere except for a delta function at the origin,

$$V(x) = \frac{\hbar^2\kappa}{2m}\delta(x)\,,$$

where κ is a coupling constant with units of inverse length. By integrating the Schrödinger equation over a small spatial interval, $-\epsilon \le x \le \epsilon$, derive an equation for the discontinuity in the first derivative of the wave function at $x = 0$. You may assume continuity of the wave function itself. Using these conditions, show that the transfer matrix relating amplitudes on the right of the delta function to those on the left is given by

$$\mathbf{M} = \begin{pmatrix} 1 - \frac{i\kappa}{2k} & -\frac{i\kappa}{2k} \\ \frac{i\kappa}{2k} & 1 + \frac{i\kappa}{2k} \end{pmatrix},$$

where the wave vector is $k = \sqrt{2mE/\hbar^2}$. Hence, determine the transmission and reflection coefficients for the delta function.

QIP

Quantum information in the solid state
Andrew Fisher

1 Introduction

1.1 Scope

These lectures will cover

- Basic introduction to the concept of a qubit and the gate model of quantum computing (QC);
- Introduction to quantum operations and quantum error correction;
- Two of the most important kinds of solid-state qubits (electron spins in semiconductors and superconducting qubits);
- Two alternatives to the standard gate model of QC which may be particularly attractive in solid-state systems: adiabatic computation and topological computation.

They will *not* cover lots of other crucial and fascinating concepts in QIP, including

- Fault-tolerant computation (except in passing in connection with topological codes);
- Quantum cryptography;
- Quantification and detection of entanglement;
- Other gate designs not based on solids (especially those involving linear optics, trapped atoms or ions);
- Measurement-based or 'cluster state' computation;
- The theory of noise and decoherence in quantum systems.

1.2 General references

Good general references for quantum information and quantum computation are

- M. Neilsen and I.L. Chuang *Quantum Computation and Quantum Information* (Cambridge 2000, ISBN 0 521 63503 9);
- J.D. Preskill lecture notes for CalTech course Ph219 (available online).

A good (reasonably modern) survey of potential implementations of quantum information processing (solid-state and otherwise) can be found in

- T.D. Ladd *et al. Nature* **464** 45 (2010), doi:10.1038/nature08812.

More specific references will be given throughout the notes.

2 The gate model

Loosely a quantum computer can be defined as a system capable of performing an arbitrary unitary transformation on a defined set of quantum states.

2.1 The Di Vincenzo Criteria

This requires

1. A well defined set of states, ideally composed of a tensor product of quantum two-level systems (qubits);
2. A means of initialising a chosen state (in principle a pure state) from the computational space;
3. Long decoherence times of arbitrarily chosen states from that set;
4. A means to implement an arbitrary unitary transformation on the chosen states;
5. An efficient readout.

To this are nowadays added (for quantum communication, or for the establishment of entangling gates between 'remote' qubits)

1. A means of interfacing static qubits faithfully with some form of mobile ('flying') qubit;
2. Faithful transmission of flying qubits between specified locations.

2.2 Universal gates

Any unitary operator can be represented by a product of unitaries acting on just two *levels*: specifically, for any unitary acting in a d-dimensional space we can find at most $d(d-1)/2$ two-level unitaries such that

$$U_k \dots U_2 U_1 U = 1 \tag{1}$$

or

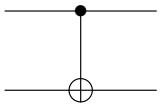
$$U = U_1^{-1} U_2^{-1} \dots U_k^{-1} = U_1^\dagger U_2^\dagger \dots U_\dagger^{-1}. \tag{2}$$

This can be done by systematically choosing each U_i to generate zeros in each off-diagonal position in the computational basis, until a final two-level unitary is reached.

Now suppose U is a two-level operator acting non-trivially only on the states $|s\rangle$ and $|t\rangle$. We show that a single two-bit gate (the CNOT) plus arbitrary single-qubit operations suffices to implement an arbitrary unitary. The CNOT (or controlled X, or controlled swap) gate flips the 'target' spin if and only if the 'control' spin is 1:

$$|c\rangle|t\rangle \to |c\rangle|t \oplus c\rangle \quad \Rightarrow \quad U = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{3}$$

Its circuit symbol is

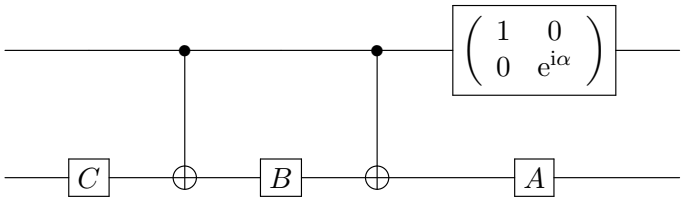


The 'plus' on the lower rail represents addition modulo 2 (could also be written as the Pauli σ_x) while the solid 'blob' on the upper rail represents the control operation ('do if and only if this is 1').

We can express an arbitrary single-qubit unitary as

$$U = \mathrm{e}^{\mathrm{i}\alpha} AXBXC \quad \text{with} \quad ABC = 1. \tag{4}$$

Single-qubit gates are written as a box representing the effecf of that unary on a single-qubit rail. Then the following circuit implements an arbitrary controlled unitary, i.e. it implements U on the target if and only if the control is 1:



This works because, if the control bit is set to 1, $U = \mathrm{e}^{\mathrm{i}\alpha} AXBXC$ is applied to the target; otherwise, $ABC = 1$ is applied.

Construct a *Gray code* connecting s and t (a sequence of binary numbers such that each differs from its successor by exactly one digit). Let the elements of the code be g_i, with $g_1 = s$ and $g_m = t$; clearly $m \leq n + 1$, since s and t differ in at most n locations.

Suppose g_1 and g_2 differ by a flip of bit i; perform a controlled swap between states of i, conditional upon the values of the other qubits being those occurring in g_1 and g_2. Now repeat until we reach a swap of g_{m-2} and g_{m-1}; then apply the one-qubit unitary at the single bit whose value is distinct in g_{m-1} and g_m, controlled so as to be conditional on the other qubits having the values taken in g_{m-1} and g_m. Finally undo the swap actions.

Hence any multi-qubit unitary can be decomposed (not necessarily efficiently) into single-qubit unitaries and controlled swaps.

In fact there are smaller sets of universal gates (i.e. not all single-qubit unitaries are required): for example, the CNOT and the $\pi/8$ phase gate are a universal set. The appropriate decompositions of single-qubit unitaries can be found efficiently using the Solovay-Kitaev theorem (see Neilsen and Chuang, also [1]).

2.3 Quantum operations and the operator-sum representation

Consider a system S interacting with an environment E, such that the total Hilbert space can be written as the product of the Hilbert spaces of each.

If we make some general unitary operation $\hat{U}$ on the system and its environment, what is its effect on the system? First suppose that the overall density operator is initially a direct product $\hat{\rho}_S \otimes \hat{\rho}_E$. (This is a significant approximation—we'll come back to it later.) Let $\{|e_k\rangle\}$ be an orthonormal basis for the environment, and let $\hat{\rho}_E = |e_0\rangle\langle e_0|$ (i.e., suppose that the environment is in the pure state $|e_0\rangle$). This sounds like a further approximation, but in fact isn't; suppose we had an environmental density operator corresponding to the mixed state

$$\hat{\rho}_E = \sum_i p_i |\psi_i\rangle\langle\psi_i|, \tag{5}$$

where the N states $\{\psi_i\}$ are not necessarily orthogonal but are normalized, and $\sum_i p_i = 1$. Then we can always introduce an additional 'far environment', F, with an orthonormal set of at least N

states $\{|f_i\rangle\}$. The following pure state of the combined $E+F$ system,

$$|\Psi\rangle = \sum_i \sqrt{p_i}|\psi_i\rangle|f_i\rangle, \tag{6}$$

has the property that its reduced density matrix in the original environment is

$$\mathrm{Tr}_F[|\Psi\rangle\langle\Psi|] = \sum_{ij} \sqrt{p_i p_j}|\psi_i\rangle\langle\psi_j|\mathrm{Tr}_F[|f_i\rangle\langle f_j|] = \sum_i p_i|\psi_i\rangle\langle\psi_i| = \hat{\rho}_E, \tag{7}$$

and it is therefore indistinguishable (as far as any measurement within E only is concerned) from the original density matrix $\hat{\rho}_E$. This is referred to as a 'purification' of $\hat{\rho}_E$. For the moment we will suppose this has been done, and the original environment E replaced by a new, bigger, environment (which we will still, however label as E) in a pure state.

Now apply $\hat{U}$ to $\hat{\rho}_S \otimes \hat{\rho}_E$ to produce the state

$$\begin{aligned} \mathcal{E}(\hat{\rho}_S) &= \mathrm{Tr}_E[\hat{U}(\hat{\rho}_S \otimes \hat{\rho}_E)\hat{U}^\dagger] & (8)\\ &= \sum_k \langle e_k|\hat{U}(\hat{\rho}_S \otimes |e_0\rangle\langle e_0|)\hat{U}^\dagger|e_k\rangle & (9)\\ &= \sum_k \hat{E}_k\hat{\rho}_S\hat{E}_k^\dagger, & (10) \end{aligned}$$

where

$$E_k \equiv \langle e_k|\hat{U}|e_0\rangle. \tag{11}$$

Note that

$$\mathrm{Tr}_S[\mathcal{E}(\hat{\rho})] = \mathrm{Tr}_S\left[\sum_k E_k\hat{\rho}E_k^\dagger\right] = \mathrm{Tr}_S\left[\sum_k E_k^\dagger E_k\hat{\rho}\right] = 1 \tag{12}$$

for any $\hat{\rho}$, so it follows that

$$\sum_k E_k^\dagger E_k = \hat{1}_S. \tag{13}$$

What sort of thing is $\mathcal{E}$? It is more general than an ordinary operator, because it acts on density *operators* of the system, not on states of it. Hence it is called a **super-operator** (Preskill) or a **quantum operation** (Neilsen and Chuang).

2.3.1 The requirements for a quantum operation

It is clear from the way $\mathcal{E}$ was introduced that any quantum operation ought generally to have certain properties.

1. It should *preserve the normalization* of the state:
$$\mathrm{Tr}[\mathcal{E}(\hat{\rho})] = 1 \quad \text{if} \quad \mathrm{Tr}\hat{\rho} = 1. \tag{14}$$
2. It should be *linear*:
$$\mathcal{E}(\sum_i p_i\hat{\rho}_i) = \sum_i p_i\mathcal{E}(\hat{\rho}_i). \tag{15}$$
3. It is *completely positive*: if we choose any possible environment E and any possible density joint density matrix $\hat{\rho}$ of the system and environment, then the result of the composite operation $(\mathcal{I}\otimes\mathcal{E})\hat{\rho}$ is another positive operator. (This requirement includes, but is more general than, the requirement that $\mathcal{E}(\hat{\rho}_S)$ be positive for any system density matrix $\hat{\rho}_S$.)

Most generally, a quantum operation is simply *defined* as a map from density operators to other density operators satisfying these conditions.

2.3.2 The Kraus representation theorem

It turns out that any quantum operation satisfying the conditions in §2.3.1 can be expressed in the form

$$\mathcal{E}(\hat{\rho}) = \sum_k \hat{E}_k \hat{\rho} \hat{E}_k^\dagger, \tag{16}$$

with

$$\sum_k E_k^\dagger E_k = \hat{1}. \tag{17}$$

The formula (16) is known as the *Kraus representation* or *operator-sum representation* of the quantum operation; the operators $\{\hat{E}_k\}$ known as the *Kraus operators*. Proof: see Preskill §3.3, or Neilsen and Chuang §8.2.4.

2.4 Introduction to quantum error correction

In practice, some level of decoherence is inevitable and limits our ability to perform any unitary operation exactly. The practical application of the circuit model therefore relies on being able to correct these errors. Basic error-correction schemes are based on quantum generalisations of classical codes which have been known for many years.

2.4.1 Classical linear codes

The *generator* of a linear code encoding k bits of information into n-bit codewords is an $n \times k$ matrix G such that the encoded form y of message x (both represented as column vectors) is simply $y = Gx$ (with multiplication and addition interpreted single-bit modulo 2). For example

$$G = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \tag{18}$$

gives the simple repetition code

$$0 \to (0,0,0); \quad 1 \to (1,1,1). \tag{19}$$

with $k = 1$, $n = 3$. Alternatively it may be specified by the $(n-k) \times n$ parity-check matrix H such that

$$Hy = 0 \quad \text{for all codewords } y \quad \Rightarrow \quad HG = 0. \tag{20}$$

Thus the columns of G must span the kernel of H (the space on which $Hy = 0$), while the rows of H are orthogonal to the columns of G. If a message is contaminated by an error e so that we receive $y' = y + e$, it follows that $Hy' = He$ (the *error syndrome*)—the syndrome is the result of a test that tells us if the message has been corrupted. For the simple repetition code above,

$$H = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \tag{21}$$

The distance d of a code is the minimum Hamming distance between any two codewords (i.e. the minimum number of flips needed to go from one to another); we then speak of an $[n, k, d]$ code. If $d \geq 2t + 1$ the code can correct t arbitrary single-bit errors (because there is still a unique 'nearest' codeword even after t bit flips); t is the *weight* of the code. Our simple repetition code has $d = 3$ and so can connect $t = 1$ error.

Important constraints include:

- $n - k \geq (d-1)$ (Singleton bound). This follows because, if the codewords are different in at least d places, we can delete $(d-1)$ digits from each of them and still have a set of distinct codewords. The maximum number of such codewords is $2^{n-(d-1)}$; hence the original number of codewords satisfies
$$2^k \leq 2^{n-(d-1)} \quad \text{or} \quad n - k \geq (d-1). \tag{22}$$
- $\frac{k}{n} \geq 1 - H(t/n)$ (Gilbert-Varshamov bound), where $H(x) \equiv -x \log_2 x - (1-x)\log_2(1-x)$ is the binary entropy function.

The *dual code* $C^\perp$ has generator H^T and parity check G^T; it encodes n bits into $n-k$ codewords, and every codeword in $C^\perp$ is orthogonal (modulo-2) to every codeword in C. A self-dual code is one where $C = C^\perp$; clearly such a code has to have $n = 2k$ so $n - k = k$.

2.4.2 Quantum codes

We wish to map our quantum states into a code space C which forms part of a larger Hilbert space. Let $\hat{P}$ be a projection operator onto the code space. The most general thing that can happen to our system is some quantum operation $\mathcal{E}$. Then for a noise operation $\mathcal{E}$ and error correction $\mathcal{R}$ we require
$$(\mathcal{R} \circ \mathcal{E})(\rho) \propto \rho \tag{23}$$
($\propto$ to allow for the possibility that the operations might not be trace-preserving). Then, in terms of the quantum operation elements (Krauss operators) $\hat{E}_k$, the condition for the existence of an $\mathcal{R}$ is that
$$\hat{P}\hat{E}_i^\dagger \hat{E}_j \hat{P} = \alpha_{ij}\hat{P} \tag{24}$$
for some Hermitiain α. The proof involves diagonalizing α and showing that under this transformation the effect of $\mathcal{E}$ is to project the system into a set of orthogonal subspaces; it can be recovered by measuring in which subspace the system lies, and devising an $\mathcal{R}$ that can correct it.

A quantum (binary) code is a map of the Hilbert space spanned by k qubits (i.e. a 2^k dimensional space) into the Hilbert space of n qubits. One simple example would be the following three-qubit 'repetition code' for encoding the state of a single qubit ($k = 1, n = 3$):
$$|0\rangle \to |0_L\rangle = |000\rangle; \qquad |1\rangle \to |1_L\rangle = |111\rangle. \tag{25}$$
This is robust against individual bit flips: if we apply the operator $\hat{X}_1$ we produce $|100\rangle$ and $|001\rangle$ respectively. But how can we tell whether there is an error, without measuring the complete state and destroying the superposition? The answer is to measure *pairs* of qubits, $\hat{Z}_1\hat{Z}_2$ and $\hat{Z}_2\hat{Z}_3$: both give +1 for the actual code space (hence do not destroy superposition of codewords), but will give a result -1 if an error has occurred. Note

- If the error has occurred with small probability, the measurement will project into either the erroneous state (with small probability) or the error-free state (with large probability).
- There is no need to measure $\hat{Z}_3\hat{Z}_1$, since it is equal to the product of the other two values.

The result of this measurement is the *error syndrome* (a classical value). Depending on the result we can take appropriate action:

- If $\hat{Z}_1\hat{Z}_2 = +1$ and $\hat{Z}_2\hat{Z}_3 = +1$ no qubit is flipped, so do nothing;
- If $\hat{Z}_1\hat{Z}_2 = -1$ but $\hat{Z}_2\hat{Z}_3 = +1$ qubit 1 is flipped, so apply $\hat{X}_1$;

- If $\hat{Z}_1\hat{Z}_2 = -1$ and $\hat{Z}_2\hat{Z}_3 = -1$ qubit 2 is flipped, so apply $\hat{X}_2$;
- If $\hat{Z}_1\hat{Z}_2 = +1$ but $\hat{Z}_2\hat{Z}_3 = -1$ qubit 3 is flipped, so apply $\hat{X}_3$.

Our code can correct single bit flips; it will fail if two bits are flipped.

So far so good - but unfortunately the code is not robust against all single-qubit errors. Consider phase errors: if we apply $\hat{Z}_1$ instead, then superpositions between the codewords are changed

$$\frac{|000\rangle + |111\rangle}{\sqrt{2}} \rightarrow \frac{|000\rangle - |111\rangle}{\sqrt{2}}; \qquad \frac{|000\rangle - |111\rangle}{\sqrt{2}} \rightarrow \frac{|000\rangle + |111\rangle}{\sqrt{2}}. \tag{26}$$

To protect against this type of error we can write the original code in the basis of X-eigenstates $|+\rangle$ and $|-\rangle$:

$$|0\rangle \rightarrow |0_L\rangle = |+++\rangle; \qquad |1\rangle \rightarrow |1_L\rangle = |---\rangle. \tag{27}$$

and measure $\hat{X}_1\hat{X}_2$ instead of $\hat{Z}_1\hat{Z}_2$ etc. This code is robust against single phase flips, but not against bit flips—so it is still not useful for general errors. (Note that $\mathrm{i}\hat{Y} = \hat{Z}\hat{X}$, so the possibility of a $\hat{Y}$ flip introduces nothing worse.)

We can achieve full protection to both types of error by (for example) expanding the code to nine qubits (the Shor code) and writing

$$\begin{aligned} |0\rangle &\rightarrow |0_L\rangle = \frac{(|000\rangle + |111\rangle)_{123}(|000\rangle + |111\rangle)_{456}(|000\rangle + |111\rangle)_{789}}{2^{3/2}} \\ |1\rangle &\rightarrow |1_L\rangle = \frac{(|000\rangle - |111\rangle)_{123}(|000\rangle - |111\rangle)_{456}(|000\rangle - |111\rangle)_{789}}{2^{3/2}} \end{aligned} \tag{28}$$

The error syndrome for this code consists in the measurement of

$$\hat{O}_1 = \hat{Z}_1\hat{Z}_2; \tag{29}$$
$$\hat{O}_2 = \hat{Z}_2\hat{Z}_3; \tag{30}$$
$$\hat{O}_3 = \hat{Z}_4\hat{Z}_5; \tag{31}$$
$$\hat{O}_4 = \hat{Z}_5\hat{Z}_6; \tag{32}$$
$$\hat{O}_5 = \hat{Z}_7\hat{Z}_8; \tag{33}$$
$$\hat{O}_6 = \hat{Z}_8\hat{Z}_9; \tag{34}$$
$$\hat{O}_7 = \hat{X}_1\hat{X}_2\hat{X}_3\hat{X}_4\hat{X}_5\hat{X}_6; \tag{35}$$
$$\hat{O}_8 = \hat{X}_4\hat{X}_5\hat{X}_6\hat{X}_7\hat{X}_8\hat{X}_9. \tag{36}$$

The first three pairs of measurements detect bit-flip errors in the first, second and third clusters of three qubits respectively. The final two detect phase flips somewhere in each set of three qubits. If any of the first set of measurements yields -1 we apply the corresponding $\hat{X}$ correction; for the second set, the pattern of -1 results tells us in which three-qubit cluster there is a phase flip—we don't need to know *where* within each cluster the phase error occurs. For example, in the first cluster, a phase flip applied to the first qubit has the effect

$$\hat{Z}_1\frac{(|000\rangle + |111\rangle)}{\sqrt{2}} = \frac{(|000\rangle - |111\rangle)}{\sqrt{2}} = \hat{Z}_2\frac{(|000\rangle + |111\rangle)}{\sqrt{2}} = \hat{Z}_3\frac{(|000\rangle + |111\rangle)}{\sqrt{2}} \tag{37}$$

so we can correct any of the three errors by applying (say) $\hat{Z}_1$.

2.4.3 Quantum stabiliser codes

Suppose a group of transformations G acts on a set X; the *stabiliser* associated with an element $x \in X$ is the subset of elements of G that fix x, i.e. it contains those elements of G for which $g.x = x$. The *normaliser* $N(S)$ of a subgroup S is the set of elements g for which $gS = Sg$, i.e. which generate similarity transformations among the elements of S.

The Pauli group on n qubits is defined as the tensor product of the groups consisting of the one-qubit Pauli matrices and ± 1, $\pm \mathrm{i}$:

$$G_1 = \{\pm 1, \pm \mathrm{i}, \pm \hat{X}, \pm \mathrm{i}\hat{X}, \pm \hat{Y}, \pm \mathrm{i}\hat{Y}, \pm \hat{Z}, \pm \mathrm{i}\hat{Z}\}. \tag{38}$$

The corresponding tensor products of Pauli operators on n qubits define the Pauli group G_n.

The term *stabiliser of a quantum state* generally refers specifically to the Pauli group; it is therefore the group of Pauli operations that leave the state invariant. We can specify the stabiliser by listing its generators. These can be conveniently summarised in a *check matrix*: if there are l generators, the matrix has l rows and $2n$ columns. The first n contain the qubits for which the generator contains an $\hat{X}$, the next n contain a 1 if the generator contains a $\hat{Z}$ for that qubit. Qubits for which there is no 1 in any column contain the identity operator for that generator. Let $r(g)$ be the row corresponding to element g; then g and g' commute if and only if

$$r(g)\Lambda[r(g')]^T = 0 \quad \text{where} \quad \Lambda = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} \tag{39}$$

(a symplectic product), and the generators are independent if and only if the rows of the matrix are independent.

We can generate a quantum code (an $[[n, k, d]]$ stabiliser code) $C(S)$ by taking the vector space V_S stabilised by a stabiliser group S having $(n-k)$ commuting generators g_r. The *normaliser* of S is then the set of elements E of G_n which map S onto itself via similarly transformations, i.e. the set such that $EgE^\dagger \in S \,\forall\, g \in S$. The normaliser has $2n - (n-k) = (n+k)$ generators. Then the correctable set of errors $\hat{E}_j$ for the code are those for which

$$\hat{E}_j^\dagger \hat{E}_k \notin N(S) - S \tag{40}$$

This follows because if (40) is satisfied either (i) $\hat{E}_j^\dagger \hat{E}_k \in S$ (in which case $\hat{P}\hat{E}_j^\dagger \hat{E}_k \hat{P} = \hat{P}$ since the projector is invariant under any element of S), or (ii) $\hat{E}_j^\dagger \hat{E}_k \in G_n - N(S)$, in which case $\hat{E}_j^\dagger \hat{E}_k$ anticommutes with at least one generator. Call this generator g_1. Then

$$\begin{aligned} \hat{P}\hat{E}_j^\dagger \hat{E}_k \hat{P} &= \prod_{l=1}^{n-k} \left(\frac{(I+g_l)}{2}\right) \hat{E}_j^\dagger \hat{E}_k \prod_{l=1}^{n-k} \left(\frac{(I+g_l)}{2}\right) \\ &= \prod_{l=2}^{n-k} \left(\frac{(I+g_l)}{2}\right) \left(\frac{(I+g_1)}{2}\right) \left(\frac{(I-g_1)}{2}\right) \hat{E}_j^\dagger \hat{E}_k \prod_{l=2}^{n-k} \left(\frac{(I+g_l)}{2}\right) \\ &= 0. \end{aligned} \tag{41}$$

We correct the code by measuring the generators $g_1 \ldots g_{n-k}$ in turn to produce an error syndrome (set of measurement outcomes β_l). The error can be located by figuring out which $\hat{E}_j$ produces the observed syndrome via the expression $\hat{E}_j g_l \hat{E}_j^\dagger = \beta_l g_l$, then corrected by applying the operator $\hat{E}_j^\dagger$.

The *weight* of an error is the number of terms in the tensor product defining an element of G_n not equal to the identity; the *distance* d is the minimum weight of any element of $N(S) - S$ (i.e.

the minimum weight of an operator which would cause an error). Equivalently, d is the minimum weight of a logical operator. A code with distance at least $d = 2t + 1$ is able to correct arbitrary errors on t qubits; it is also able to correct 'located errors' on up to $d - 1$ qubits, as long as we are told which ones are affected. The Shor code discussed above has $n = 9$, $k = 1$, $d = 3$ and $t = 1$. Its check matrix is

$$H = \left(\begin{array}{ccccccccc|ccccccccc} 0&0&0&0&0&0&0&0&0&1&1&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&1&1&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1\\ 1&1&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&1&1&1&1&1&1&0&0&0&0&0&0&0&0&0 \end{array}\right) \tag{42}$$

This is an example of a Shor-Caldebank-Steane (CSS) code: such codes combine two classical codes C_X and C_Z to form a single quantum code and have parity check matrices of the general form

$$H = \left(\begin{array}{cc} H_Z & 0 \\ 0 & H_X \end{array}\right), \tag{43}$$

so the syndrome measurements require products of X operators or Z operators, but never both. For the X generators and Z generators to commute we must have

$$H_X H_Z^T = H_Z H_X^T = 0, \tag{44}$$

so the dual of the classical code with parity check H_X must be contained in the Z-code ($C_X^\perp \subset C_Z$), and vice versa ($C_Z^\perp \subset C_X$).

The quantum Singleton bound states that $n - k \geq 2(d - 1)$ (note the factor of two on the RHS, different from the classical case). Hence $n - k \geq 4t$, so the minimum length for a code which will encode a single qubit ($k = 1$) and protect against arbitrary single-qubit errors ($t = 1$) is $n = 5$.

The $[[5, 1, 3]]$ code saturating this bound (the Reed-Muller code) is constructed as follows. The parity check matrix is

$$H = \left(\begin{array}{ccccc|ccccc} 0&1&1&0&0&1&0&0&1&0\\ 0&0&1&1&0&0&1&0&0&1\\ 0&0&0&1&1&1&0&1&0&0\\ 1&0&0&0&1&0&1&0&1&0 \end{array}\right), \tag{45}$$

so the generators of the stabiliser are

$$\hat{O}_1 = \hat{X}_1\hat{Z}_2\hat{Z}_3\hat{X}_4\hat{I}_5; \tag{46}$$

$$\hat{O}_2 = \hat{I}_1\hat{X}_2\hat{Z}_3\hat{Z}_4\hat{X}_5; \tag{47}$$

$$\hat{O}_3 = \hat{X}_1\hat{I}_2\hat{X}_3\hat{Z}_4\hat{Z}_5; \tag{48}$$

$$\hat{O}_4 = \hat{Z}_1\hat{X}_2\hat{I}_3\hat{X}_4\hat{Z}_5. \tag{49}$$

In other words, each generator contains two X operators and two Z operators, with cyclic permutations. (The fifth cyclic permutation is simply the product of the other four.) It has distance $d = 3$ because each product of one or two Pauli matrices anti commutes with at least one generator and hence lies outside $N(S)$; for example, $\hat{X}_1\hat{O}_4 = -\hat{O}_4\hat{X}_1$ and $\hat{X}_1\hat{Z}_2\hat{O}_2 = -\hat{O}_2\hat{X}_1\hat{Z}_2$. (Clearly this code is not a CSS code.) The columns of H tell us the result of measuring each stabiliser generator

if different types of error occur: for example, an X error (bit flip) on the first qubit will give -1 when $\hat{O}_4$ is measured (since it contains $\hat{Z}_1$), but $+1$ when the other generators are measured, and conversely a Z error (phase flip) on the first qubit will give a -1 result under $\hat{O}_1$ or $\hat{O}_3$. The results of measuring the check operators (the syndrome) give a unique diagnosis of the error so the code is said to be *non-degenerate*.

To manipulate our logical qubits we need to find operators which transform our logical qubits into one another—i.e. which take a state in the code space and produce another state in the code space. These are precisely the same operators that the code cannot correct; to put it another way, they commute with the stabiliser but lie outside it, hence lie $N(S) - S$. We expect that space of operators to have $2k$ generators; in this case $k = 1$ and let's choose for those two generators

$$\tilde{Z} = \hat{Z}_1\hat{Z}_2\hat{Z}_3\hat{Z}_4\hat{Z}_5 \quad \text{and} \quad \tilde{X} = \hat{X}_1\hat{X}_2\hat{X}_3\hat{X}_4\hat{X}_5. \tag{50}$$

It's easy enough to find eigenstates of $\tilde{Z}$—anything in the computational basis containing odd or even numbers of 1s will do. To pick out the correct linear combination of these states we need to find one that is invariant under all the stabiliser operations; this can be done by writing

$$|0_L\rangle = \frac{1}{4}\left[|00000\rangle + (|10010\rangle + \text{cycles}) - (|11110\rangle + \text{cycles}) - (|01100\rangle + \text{cycles})\right]; \tag{51}$$

$$|1_L\rangle = \frac{1}{4}\left[|11111\rangle + (|01101\rangle + \text{cycles}) - (|00001\rangle + \text{cycles}) - (|10011\rangle + \text{cycles})\right]. \tag{52}$$

2.5 Superconducting circuits

2.5.1 Review of Ginzburg-Landau theory

At a phenomenological level a superconductor is described by Landau-Ginzburg theory: the variational condition is

$$\alpha\psi + \beta|\psi|^2\psi + \frac{1}{2m^*}(-\mathrm{i}\hbar\nabla - e^*\mathbf{A})^2\psi = 0, \qquad \psi = |\psi|\mathrm{e}^{\mathrm{i}\phi} \tag{53}$$

and the velocity is

$$m^*\mathbf{v}_s = \hbar\nabla\phi - e^*\mathbf{A}, \tag{54}$$

with the associated current density

$$\mathbf{j} = \frac{e^*\hbar}{2m^*\mathrm{i}}\left[\psi^*\nabla\psi - \psi\nabla\psi^*\right] - \frac{(e^*)^2}{m^*}|\psi|^2A^2 = e^*|\psi|^2\mathbf{v}_s. \tag{55}$$

The effective charge $e*$ is the charge on a Cooper pair: $e^* = -2|e|$.

In one dimension with $\mathbf{A} = 0$ we get

$$\frac{\hbar^2}{2m^*|\alpha|}\frac{\mathrm{d}^2f}{\mathrm{d}x^2} + f - f^3 = 0 \quad \text{with} \quad f = \frac{\psi}{\psi_\infty}, \quad |\psi_\infty|^2 = -\alpha/\beta, \tag{56}$$

or

$$\frac{1}{\xi^2}\frac{\mathrm{d}^2f}{\mathrm{d}x^2} + f - f^3 = 0 \quad \text{with} \quad \xi^2 = \frac{2m^*|\alpha|}{\hbar^2}. \tag{57}$$

2.5.2 Josephson junctions

Consider a junction between two superconducting regions with a phase difference $\Delta\phi$. If the junction length L is small compared with ξ then the first term dominates and

$$\frac{\mathrm{d}^2 f}{\mathrm{d}x^2} = 0. \tag{58}$$

The solution to this that satisfies the given boundary conditions of $f(0) = 1$ and $f(L) = \mathrm{e}^{\mathrm{i}\Delta\phi}$ is

$$f(x) = \left(1 - \frac{x}{L}\right) + \frac{x}{L}\mathrm{e}^{\mathrm{i}\Delta\phi}. \tag{59}$$

Hence the current density is

$$\begin{aligned} j &= \frac{e^*\hbar|\psi_\infty|^2}{2\mathrm{i}m^*}\left[\left(\left(1-\frac{x}{L}\right)+\frac{x}{L}\mathrm{e}^{-\mathrm{i}\Delta\phi}\right)\left(\mathrm{e}^{\mathrm{i}\Delta\phi}-1\right)/L - \left(\left(1-\frac{x}{L}\right)+\frac{x}{L}\mathrm{e}^{\mathrm{i}\Delta\phi}\right)\left(\mathrm{e}^{-\mathrm{i}\Delta\phi}-1\right)/L\right] \\ &= \frac{e^*\hbar|\psi_\infty|^2}{m^*L}\left[\sin(\Delta\phi)\right] \end{aligned} \tag{60}$$

and for a junction of area A,

$$i(t) = I_0 \sin[\Delta\phi] \tag{61}$$

where I_0 is a constant, $\Phi = \int_{-\infty}^{t} v(t')\,\mathrm{d}t'$ is the flux (the time integral of the voltage). This is related to the conventional definition of a flux because

$$\nabla \times \mathbf{E} = -\dot{\mathbf{B}} \tag{62}$$

so by Stokes' theorem the voltage around a loop is

$$v(t) = \int_C \mathbf{E}(t)\cdot\mathrm{d}\mathbf{r} = -\int \dot{\mathbf{B}}(t)\cdot\mathrm{d}\mathbf{A} = -\dot{\Phi} \tag{63}$$

2.5.3 Circuit Hamiltonians and the RCSJ model

The Josephson current gives rise to a natural energy scale: the work done is

$$E = -E_J\cos(\Delta\phi) \quad \text{where} \quad E_J = \frac{\hbar I_0}{e^*} \tag{64}$$

So far $\Delta\phi$ has the appearance of a classical quantity (a boundary condition on the pair wave function). But we can turn in into a proper quantum mechanical operator by remembering that ψ is really the annihilation operator for a Cooper pair and thinking about the basis of states $|N\rangle$ in which *exactly* N pairs have crossed the junction; there is therefore a number operator

$$\hat{N} = \sum_N N|N\rangle\langle N| \tag{65}$$

and a dual (phase) basis describing the phase relationship when one pair is transferred across the junction

$$|\phi\rangle = \sum_N \mathrm{e}^{\mathrm{i}N\phi}|N\rangle, \tag{66}$$

such that

$$|N\rangle = \frac{1}{2\pi}\int_0^{2\pi}\mathrm{d}\phi\,\mathrm{e}^{-\mathrm{i}N\phi}|\phi\rangle \tag{67}$$

and

$$\widehat{\mathrm{e}^{i\phi}} = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\phi\, \mathrm{e}^{\mathrm{i}\phi}|\phi\rangle\langle\phi| \tag{68}$$

satisfies

$$\widehat{\mathrm{e}^{i\phi}}|N\rangle = |N-1\rangle. \tag{69}$$

The relationship between number and phase is entirely analogous to that between momentum and position, so it follows that

$$[\hat{\phi}, \hat{N}] = \mathrm{i}, \tag{70}$$

and we can represent

$$\hat{N} = -\frac{\partial}{\partial\phi}. \tag{71}$$

Written in terms of number eigenstates, the junction Hamiltonian take the form

$$\hat{H}_J = -E_J\widehat{\cos(\phi)} = -\frac{E_J}{2}\sum_N \left[|N\rangle\langle N-1| + |N-1\rangle\langle N|\right]; \tag{72}$$

in other words E_J is the tunnelling energy connecting states of different N (it is like a tight-binding hopping element connecting different atoms). It can also be thought of as a characteristic energy resisting the imposition of a phase difference across the junction: the more transparent the junction, the larger is E_J and the more strongly the junction resists any phase difference across it that is not a multiple of 2π.

In reality the junction will always be found in parallel with a capacitance describing the interaction of the Cooper pairs on the two sides:

$$\hat{H}_C = \frac{(e^*)^2(\hat{N}-N_g)^2}{2C} \tag{73}$$

where $N_g = V_g/(e^*C_g)$ is the number of pairs required to make the junction neutral under the electrostatic action of the gate potential. If there is also an inductance L in the circuit there will be an associated inductive term

$$\hat{H}_L = \frac{\hat{\Phi}^2}{2L} \tag{74}$$

where the magnetic flux Φ is related to the superconducting phase ϕ by

$$\Phi = \frac{\hbar}{e^*}\phi \tag{75}$$

The 'flux quantum' h/e^* is thus the magnetic flux associated with a superconducting phase of 2π.

A linear LC circuit containing just a capacitor and an inductor in series therefore has the Hamiltonian (in the absence of any external gate bias)

$$\hat{H} = \hat{H}_C + \hat{H}_L = -\frac{(e^*)^2}{2C}\frac{\partial^2}{\partial\phi^2} + \frac{\hbar^2}{2(e^*)^2L}\phi^2 = -\frac{\hbar^2}{2C}\frac{\partial^2}{\partial\Phi^2} + \frac{1}{2L}\Phi^2. \tag{76}$$

This is nothing but a simple harmonic oscillator with angular frequency $\omega^2 = (LC)^{-1}$, and corresponding energy eigenvalues $E_n = (n+\frac{1}{2})\hbar\omega$.

If there is also an external current source I_S in the circuit, it can be represented as a large inductance L_S with an initially large (classical) magnetic flux Φ_S such that

$$\phi_S = I_S L_S. \tag{77}$$

Then the total Hamiltonian contains an inductive term like $(\Phi_S - \Phi)^2/(2L_S)$, which up to an additive constant is equivalent to

$$\hat{H} = \hat{H}_C + \hat{H}_L + \hat{H}_J - \frac{e^*}{\hbar} I_S \hat{\phi}. \tag{78}$$

This corresponds to a 'particle' of mass C moving in a 'potential'

$$V(\phi) = -\frac{e^*}{\hbar} I_S \hat{\phi} - E_J \widehat{\cos(\phi)} + \frac{\hat{\Phi}^2}{2L}. \tag{79}$$

This is known as the 'tilted washboard' potential; it contains a slowly varying (parabolic) term plus an oscillating part.

2.5.4 Josephson (flux and phase) qubits

For circuits dominated by the Josephson term ($E_J \gg E_C$), the oscillating 'potential' energy is much more important than the 'kinetic' energy when the landscape is viewed as a function of ϕ. Now it matters how the junction is biased: if the slowly varying part $-I_S\Phi + \hat{H}_L$ is chosen to pick out a single 'well' lower than all the others, then the circuit is strongly asymmetric and the lowest two states are both found in that well. These two states (with similar fluxes but different wavefunctions) can be used as the basis states states for a qubit; this is known as a 'phase qubit'. Alternatively, if the slowly varying part is a minimum exactly half-way between two wells then the two lowest states are the symmetric and anti-symmetric combinations of the lowest levels of the left and right wells; this is a 'flux qubit'.

2.5.5 Superconducting charge qubits

If the charging term dominates the Josephson term, $\hat{N}$ is 'almost' a good quantum number. This situation is sometimes referred to as a 'Cooper pair box' [2]: suitable computational basis states then correspond to N and $(N+1)$ (with a value of N picked out by the gate bias potential V_g). A particularly clever choice is to make the Hamiltonian of the circuit insensitive to both charge and magnetic fluctuations; this can be accomplished by choosing E_J and E_C so that they are comparable (hence neither $\hat{N}$ nor $\hat{\phi}$ is a good quantum number), and choosing V_g so that $N_g = N + \frac{1}{2}$. The lowest two states then have the same values of both $\langle \hat{N} \rangle$ and $\langle \hat{\phi} \rangle$. This is known as a 'quantronium' or 'transmon' qubit [3]. The states cannot be manipulated (to first order) by changing the dc bias, but they can be coupled by an oscillating electromagnetic field (typically a microwave field) in the same manner as the states of a spin qubit.

2.5.6 Selection of recent progress

- Single-shot readout, with the introduction of an external shunting capacitance [4];
- Multi-qubit gates [5];
- Coupling of superconducting qubits to harmonic oscillators ('transmission lines', forming flying qubits) and the corresponding start of cavity QED for superconductors [6].
- Implementation of simple error correction (three-qubit bit-flip and phase-flip codes) [7];
- Coupling of 'processing' to 'memory' qubits [8] and the execution of simple algorithms including factoring [9].

- Large-scale planar architectures [10] and coupling of neighbouring qubits to produce two-qubit fidelities above thresholds for fault-tolerant operation [11]

2.6 Spins in semiconductors

Spins with $S = 1/2$ form a natural two-level system, the states $|0\rangle$ and $|1\rangle$ corresponding to $|S_z = -1/2\rangle$ and $|S_z = +1/2\rangle$ respectively, with magnetic moments $\mp\mu_B$. Their states can be manipulated by external microwave fields as follows: apply a static magnetic field B_z and an oscillating magnetic field in the xy plane, inducing transitions between the spin states: this corresponds to a variant of the well-known technique of electron paramagnetic resonance (EPR).

2.6.1 Spins at donors

One natural way to localise spins is to bind them at impurity sites in a solid. The best studied examples are (single) donors in semiconductors: these have hydrogenic states whose energy scale is reduced from the free-space Rydberg by a factor $\frac{m^*}{m_e}\left(\frac{\epsilon_0}{\epsilon}\right)^2$ (typically in the meV range). P (binding energy 46 meV) and Bi (binding energy 71 meV) [12] are two common choices. These can be manipulated in bulk using microwave magnetic fields to produce transitions between the spin states; they can also be coupled to the nuclear spin using the hyperfine interaction. Very long coherence times have been demonstrated, exploiting the greater robustness of the deeper donors and/or the absence of other decohering nuclear spins in isotopically pure ^{28}Si [13, 14]. Most of these have been ensemble measurements in a temperature regime where the thermal state of the donors is only weakly perturbed; recent developments include the detection of the spins state of single impurities [15], the manipulation of spin in cooled devices so as to produce genuine entanglement [16], and very recently [17] the combination of these to perform and detect spin manipulations at a single atomic site. It has not so far been possible to connect donors to perform multii-qubit operations outside a single donor site.

2.6.2 Optical control of defect spins

Here the favoured system is the NV-centre in diamond, which can be conveniently initialised optically and have reasonably long coherence times even at room temperature. Here there is a convenient (visible) optical transition which can be used to polarise the defect in its ground (triplet) state and to detect its spin state, and microwaves can then be used to manipulate the system within the ground-state manifold [18]. Multi-qubit experiments have been performed using the nuclei of nearby ^{13}C atoms to form a simple quantum register [19]. A simple form of quantum error correction has been demonstrated using this system [20]. The most likely prospect for larger-scale coupling of different NV centres is optically.

An alternative route for optical control has been proposed through control of the extent of donor electron wave functions [21]; this requires working in the experimentally inconvenient far-infrared or THz region of the spectrum. Coherent control of the orbital state in an ensemble has been demonstrated, but not yet control of the spins [22].

2.6.3 Gate control of spins in quantum dots

The alternative is to use an artificial atom or quantum dot. These can be created in designed locations using lithographically-defined gate electrodes, and the couplings between them controlled using the gate voltages. This enables, for example, manipulation of the singlet and triplet states of two excess electrons in a double-dot structure [23] or of a single spin in an isolated dot [24].

Unfortunately the easiest material from an electronic point of view for these experiments is GaAs, which suffers from a very high density of nuclear spins, necessitating special techniques to prepare the spin bath in order to attain reasonable coherence times [25].

2.6.4 Optical control of spins in quantum dots

Alternatively, working with smaller dots grown into the chemical structure of the material, one can read out and manipulate the spin of a charged excited state (e.g. the so-called X^- state) much more quickly using optical means [26].

3 Adiabatic quantum computing

3.1 Simulated annealing and quantum annealing

3.1.1 Experiments

Pioneered in LiHoF4. DiIute random (nearly Ising) ferromagnet: effective Hamiltonian of the form

$$\hat{H} = h \sum_i \hat{\sigma}_i^z + \Delta \sum_i \hat{\sigma_i}^x + \sum i, j J_{ij} \hat{\sigma}_i^z \hat{\sigma}_j^z, \tag{80}$$

where h and Δ are (globally) controllable, J_{ij} are random and fixed by the microscopic disorder. The original experiments contrasted reaching a given state through different paths in the (Δ, T) plane and found increased effectiveness in locating the low-temperature state when transverse-field ('quantum') terms were included [27].Santoro

3.1.2 Annealing algorithms—successes and failures

One of the most effective ways to solve hard optimisation problems is via simulated annealing. A Hamiltonian is constructed corresponding to the cost function to be minimised, and the temperature is gradually reduced. Provided the temperature reduction is sufficiently slow, the system will end up in the configuration with the lowest cost function with high probability.

Simulations contrasted purely 'classical' Monte Carlo with a form of 'quantum' dynamics generated by a path-integral Monte Carlo. (This is not a true simulation of a noisy quantum system but does correspond to one way of incorporating quantum effects in a classical simulation.) In certain circumstances the addition of a 'quantum' (i.e. non-commuting) contribution to the Hamiltonian can speed the attainment of a ground state corresponding to the solution of a 'hard' problem: this worked for random instances of the travelling salesman problem, but not for random instances of the 3–SAT circuit satisfiability problem [28, 29].

3.2 Ideal adiabatic computation

3.2.1 The adiabatic theorem

Consider a system whose Hamiltonian is slowly changed from an initial $\hat{H}_0$ to a final $\hat{H}_f$. Quantify this by writing

$$\hat{H}(s=0) = \hat{H}_0; \quad \hat{H}(s=1) \tag{81}$$

and put $s = t/\tau$ where τ is a constant (the 'duration' of the evolution). Suppose $E_0(s)$ is the trajectory of an eigenvalue of $\hat{H}$, which we suppose can be continuously defined and always non-degenerate; let the minimum gap between this eigenvalue and any other eigenvalue be $\gamma(s)$. The

time-evolution operator $\hat{U}(s)$ satisfies

$$\mathrm{i}\frac{\mathrm{d}\hat{U}}{\mathrm{d}s} = \tau\hat{H}(s)\hat{U}(s). \tag{82}$$

Let $\hat{P}_0$ be the projector on to the eigenvector of $\hat{H}_0$ corresponding to $E_0(0)$, and $\hat{P}$ be the corresponding projection at some later value of s. Then [30] the matrix elements of $\hat{U}$ taking us outside the subspace corresponding to the chosen E_0 at some particular s are bounded by

$$||(1-\hat{P})\hat{U}\hat{P}_0|| \leq c(k)\left(\frac{1}{\tau}\frac{1}{\gamma^2}h\bigg|_0 + \frac{1}{\tau}\frac{1}{\gamma^2}h\bigg|_s + \max_{0\leq\sigma<s}\frac{1}{\tau^k}\frac{1}{\gamma^{2k+1}}h^{k+1}\right), \tag{83}$$

where $c(k)$ is c constant, the Hamiltonian is assumed to be $(k+1)$-times differentiable and its derivatives are bounded by

$$||\left(\frac{\mathrm{d}}{\mathrm{d}s}\right)^l\hat{H}(s)|| \leq h(s) \quad \forall l \leq k. \tag{84}$$

Clearly, by choosing

$$\tau^k \gg \frac{h^{k+1}}{\gamma^{2k+1}} \quad \Rightarrow \quad \tau \gg \frac{h^{1+1/k}}{\gamma^{2+1/k}} \tag{85}$$

we can make the bound (83) as small as desired. For a perfectly smooth (infinitely-many times differentiable) Hamiltonian, $k \to \infty$ and we require

$$\tau \gg \frac{h}{\gamma^2}. \tag{86}$$

3.2.2 The algorithm

Choose a function f defined on n-bit binary numbers z. Define the Hamiltonian

$$\hat{H}_\mathrm{f} = \sum_{z\in\{0,1\}^n} f(z)|n\rangle\langle n|, \tag{87}$$

so its eigenvalues are the values of f for the binary numbers z. Its lowest eigenvalue is therefore the minimum value of f.

Choose an initial Hamiltonian that is 'easy' to prepare (typically one whose ground state is a product state), does not commute with $\hat{H}_\mathrm{f}$ and which does not share any symmetries with it. A typical choice is

$$\hat{H}_0 = \sum_z h(z)|\tilde{z}\rangle\langle\tilde{z}| \tag{88}$$

where $|\tilde{z}\rangle$ is the Hadamard transform

$$|\tilde{z}\rangle = \hat{W}^{\otimes n}|z\rangle \tag{89}$$

with

$$\hat{W}|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle); \quad \hat{W}|1\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle); \tag{90}$$

Vary the Hamiltonian slowly from $\hat{H}_0$ to $\hat{H}_f$, for example by writing

$$\hat{H}(s) = (1-s)\hat{H}_0 + s\hat{H}_\mathrm{f}. \tag{91}$$

Then the ground state of $\hat{H}_0$ (corresponding to the minimum of $h(z)$) is transformed into the minimum of $f(z)$ after a time τ [31]. The resources needed will depend on the minimum gap γ encountered during the evolution—this is difficult to determine in general.

3.2.3 Adiabatic QC is contained within gate QC

We can simulate this evolution efficiently given only a transformation to go between the eigenstates of the initial Hamiltonian and the final Hamiltonian [32]. Such a transformation is provided by the n-bit Hadamard gate:

$$\hat{W}^{\otimes n}|000\ldots0\rangle = |+++\ldots+\rangle; \quad \hat{W}^{\otimes n}|111\ldots1\rangle = |---\ldots-\rangle \tag{92}$$

so

$$\hat{W}^{\otimes n}|z\rangle = |\tilde{z}\rangle. \tag{93}$$

It is easy to apply $\hat{H}_0$ in the Hadamard basis $\{|+\rangle, |-\rangle\}$ and it is easy to apply $\hat{H}_{\rm f}$ in the computational basis $\{|0\rangle, |1\rangle\}$. The Baker-Hausdorff theorem tells us that we can simulate the action of two Hamiltonians acting 'together' (in parallel) for a short time by applying the same operators 'sequentially':

$$\mathrm{e}^{\hat{A}+\hat{B}} = \mathrm{e}^{\hat{A}}\mathrm{e}^{\hat{B}} + \mathrm{O}([\hat{A}, \hat{B}]). \tag{94}$$

So we simulate the complete adiabatic algorithm by applying the sequence $\hat{U}r\ldots\hat{U}_1$, where

$$\hat{U}_j = \hat{W}^{\otimes n}\hat{F}_{0,j}\hat{W}^{\otimes n}\hat{F}_{f,j} \tag{95}$$

and

$$\hat{F}_{0,j}|z\rangle = \mathrm{e}^{-\mathrm{i}\frac{T}{r}(1-\frac{j}{r})h(z)}|z\rangle; \qquad \hat{F}_{f,j}|z\rangle = \mathrm{e}^{-\mathrm{i}\frac{T}{r}\frac{j}{r}f(z)}|z\rangle \tag{96}$$

We need $r \sim T^2 n^{d+1}$ to obtain a constant accuracy in the simulation; this grows polynomially in n and T, hence the simulation is efficient.

3.2.4 Adiabatic QC is equivalent to gate QC

For this we need to show an adiabatic QC can simulate an arbitrary gate-based computation. This was shown in [33] using a trick of Kitaev's [34]: if the gate computation on n qubits starts WLOG in the state $|0\rangle^n$ and has the state $|\alpha(l)\rangle$ after l steps, we construct the state

$$|\eta\rangle = \frac{1}{\sqrt{L+1}}\sum_{l=0}^{L}|\alpha(l)\rangle \otimes \left|1^l 0^{L-l}\right\rangle^c = \frac{1}{\sqrt{L+1}}\sum_{l=0}^{L}|\gamma(l)\rangle, \tag{97}$$

where c is a set of 'clock' qubits labelling the steps of the calculation. Choose

$$\hat{H}_0 = \hat{H}_{\rm initclock} + \hat{H}_{\rm input} + \hat{H}_{\rm clock} \tag{98}$$

and

$$\hat{H}_{\rm f} = \sum_l \hat{H}_l + \hat{H}_{\rm input} + \hat{H}_{\rm clock}, \tag{99}$$

so $\hat{H}_{\rm initclock}$ is replaced by $\sum_l \hat{H}_l$ as the calculation proceeds.

Here

$$\hat{H}_{\rm clock} = \sum_l |01\rangle\langle 01|^c_{l,l+1} \tag{100}$$

penalises clock states containing 01 and hence enforces clock states of the type $\left|1^l 0^{L-l}\right\rangle$,

$$\hat{H}_{\rm input} = \sum_i |1\rangle\langle 1|_i \otimes |0\rangle\langle 0|^c_1 \tag{101}$$

enforces input states of the form $|0\rangle^n$ at timestep 0, while

$$\hat{H}_{\text{initclock}} = |1\rangle\langle 1|_1^c \tag{102}$$

enforces a clock state 0 in the first clock qubit. Hence the ground state of $\hat{H}_0$ is $|\gamma(0)\rangle$. The Hamiltonian $\sum_l \hat{H}_l$ enforces the correct evolution of the gate-model state at timestep l:

$$\hat{H}_l = |10\rangle\langle 10| - \hat{U}_l|11\rangle\langle 10| - \hat{U}_l^\dagger|10\rangle\langle 11| + |11\rangle\langle 11|, \tag{103}$$

The ground state of the Hamiltonian $\hat{H}_{\text{f}}$ is then given by

$$\hat{H}_I|\eta\rangle = 0. \tag{104}$$

Now we have to worry about the execution time of the adiabatic algorithm. The subspace $\{|\gamma_0\rangle, \ldots, |\gamma_N\rangle\}$ is closed under the operation of $\hat{H}$. Within this space, we have the matrix representations

$$\hat{H}_0 = \begin{pmatrix} 0 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix} \tag{105}$$

and

$$\hat{H}_{\text{f}} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 0 & \ldots & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & \ddots & \vdots \\ 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \tag{106}$$

The gap of $\hat{H}_0$ is 1; $\hat{H}_{\text{f}}$ is a discretization of the one-dimensional Laplacian (hence the 1D kinetic energy) so its gap scales as $\mathrm{O}(L^{-2})$. In fact one can show that the gap of $\hat{H}(s) = (1-s)\hat{H}_0 + s\hat{H}_{\text{f}}$ is at least $\mathrm{O}(L^{-2})$ for all $0 \leq s \leq 1$. This establishes that for any fixed $\delta > 0$, an adiabatic QC outputs a final state within distanceϵ of the desired history state η in a time $T = \mathrm{O}(\epsilon^{-\delta} L^{4+2\delta})$.

Hence adiabatic quantum computation is equivalent to standard gate-based quantum computation. The remaining question is whether adiabatic computation is a more practically useful way of thinking about computation, for example from the point of view of resistance to noise.

3.3 Adiabatic computation in the presence of noise

3.3.1 The formal position

The complexity class of problems accessible to a noisy Turing machine (i.e. one endowed with the ability to flip coins at random) is BPP (bounded-error, probabilistic, polynomial time). We have

> P is the class of all problems solvable by a deterministic Turing machine bounded in time by a polynomial function of the input length,

and

> BPP is the class of languages which can be solved in polynomial time by a P machine with access to a fair coin, with the constraint that the algorithm's error rate is bounded. By that, we mean that a BPP machine given a "yes" instance must return "yes" at least 2/3 of the time, and a BPP machine given a "no" instance must return "no" more than 1/3 of the time.

The success probability is arbitrary provided it exceeds 1/2. Most complexity theorists believe P=BPP (i.e. the probabilistic capability makes no difference to the classification of the algorithms), but no proof is known.

The complexity class of problems efficiently soluble on a quantum computer is BQP:

> BQP is the class of languages decidable by a uniform family of quantum circuits (that is, such that each circuit is produced by a P machine) involving no more than a polynomial number of gates drawn from a specified set of universal quantum gates (such as the Toffoli and Hadamard gates, or the CNOT and $\pi/8$ gates), subject to the same bounded-error constraints as BPP (we conventionally require a success probability of 2/3) [35].

From the result above, the class of problems efficiently solved by AQC is also BQP. It is known that

$$\mathrm{BQP} \supseteq \mathrm{BPP} \supseteq \mathrm{P}. \tag{107}$$

The second inequality is obvious. The first comes from noting that if a language is in BPP, there exists a polynomial-time Turing machine (M) that decides the language with the use of some random bits. Hence if we make the table of results over possible results y of the random bits accessible to a BPP machine, by computing M on the inputs (x, y) (where x is of length n and y is of polynomial length $p(n)$), the fraction of 1s in the table is guaranteed to be at least 2/3 if $x \in \mathcal{L}$, and at most 1/3 otherwise. We now construct a quantum Turing machine that averages over all elements of this table by producing the superposition

$$\frac{1}{\sqrt{2^{p(n)}}} \sum_y |x\rangle|y\rangle|M(x,y)\rangle. \tag{108}$$

Measuring the final qubit tells us whether $x \in \mathcal{L}$ with the required success probability. Hence any language in BPP is also decidable by a BQP machine in polynomial time.

So—for a classical machine the presence of noise (random bits) may or may not help. But the introduction of quantum mechanics seems very likely to help more. In other words the conjectured relationship is

$$\mathrm{BQP} \supset \mathrm{BPP} = \mathrm{P} \quad \text{(conjectured)} \tag{109}$$

Is it possible that introducing environmental noise into a quantum computer helps some more? This too is formally open: the general trace-preserving operation induced on an n-qudit system (having dimension $d(n) = d^n$) by the action of a Hamiltonian on the system and its environment is

$$\mathcal{E}(\rho) = \sum_k \hat{E}_k \hat{\rho} \hat{E}_k^\dagger. \tag{110}$$

The resulting mixed state (since it has rank at most $d(n)$), can always be purified by including an environment of at most n qudits. However there is no guarantee that the the resulting unitary evolution can be effiiciently approximated by local gates (the number of gates required to approximate a general n-qudit unitary grows exponentially with n). So it is possible that noise from a general quantum environment enables a general quantum machine to reach states in a polynomial number of steps which would require an exponential number of steps with $2n$ perfectly isolated qudits.

Note also that it is strongly suspected (but not proved) that NPC (the class of NP-complete problems) is disjoint from BQP (but of course lies within NP). These conjectures are summarised in Figure 1.

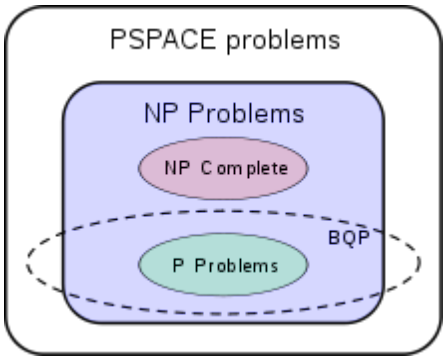


Figure 1: Conjectured relationships among complexity classes.

3.3.2 Results for a complex system

For many qubits, there are conflicting results.

- Dependence of gap on 'difficulty' of problem: it is known that 'awkward' choices of certain problems have exponentially small gaps and so cannot be efficiently simulated by AQC [32]. It is also known that 'typical' (i.e. randomly generated) instances of certain problems (for example 'exact cover 3') can have exponentially small gaps if the path through Hamiltonian space is found in the most obvious way (moving linearly from initial to final Hamiltonian) [36]. However in these cases there are deformations of the path which retain the polynomial scaling of the gap [37]; it is not known whether these deformations can in general be found efficiently.

- It is not not clear how, for realistic noise models, coupling from the environment to the so-called computational subspace (the lowest two states at the anticrossing) influences the computation [38].

3.4 The D-Wave One as a potential implementation of an annealing machine

- The system is based on superconducting flux qubits.

- The qubits are not optimized for coherence, so have short decoherence times (20 ns?)

- The system has an unusual architecture. Units containing eight qubits are inductively coupled, using an intermediate Josephson junction which enables control of the magnitude and sign of the coupling. They also have independent control of the qubit bias and transverse field, giving a Hamiltonian of the form

$$\hat{H} = \sum_i [h_i \hat{\sigma}_i^z + \Delta_i \hat{\sigma_i}^x] + \sum_{i,j} J_{ij} \hat{\sigma}_i^z \hat{\sigma}_j^z. \tag{111}$$

- Within each cluster of eight qubits there is full control of all parameters $\{h, \Delta, J\}$; between clusters, only selective coupling of the qubits is possible.

- There is some evidence for quantum behaviour in the solution of an annealing problem [39]. Specifically, the transverse field terms Δ were reduced (thus 'freeing out' fluctuations in the Ising magnetization) and it was tested for how late in the computation the system remained sensitive to the changes in the Ising part of the Hamiltonian. It was found that the 'freeze out time' saturated at low temperatures, suggesting that the fluctuations overcoming freeze-out had a quantum-mechanical origin.

3.5 Open questions (as far as I know)

1. Is quantum mechanics playing a role in the operation of the device?

2. How large a window of coherent evolutions is needed for a useful computation?

3. Are there any generic statements that can be made about the behaviour of the gap and which problems are 'hard' or 'difficult' for adiabatic computation?

4 Topological quantum computation

4.1 Why would topological states be useful for quantum computation ?

Topological states can be protected against the action of many (or all) local operators. Many noise processes involve local operators — these states can be protected against such noise. In addition, if the manipulation of the states involved only topological properties of that manipulation process, the manipulation will be robust against imprecision in the process.

4.2 The toric code and the surface code

These are topological codes introduced by Kitaev [40]. An excellent and readable summary can be found in [41]. Consider an $l \times l$ square lattice, with a spin 1/2 on the edge link between each vertex pair. Evidently there are four spins per elementary square, each being shared between two squares, so there are $2l^2$ spins. (Equivalently, if the original square lattice has lattice parameter 1, the spins form a square lattice with lattice parameter $1/\sqrt{2}$.) For the toric code we impose periodic boundary conditions in both directions; the related surface codes are considered in the problem sheet.

4.2.1 The Hamiltonian

Assign to each vertex on the original lattice a 'star' label s, and to each square unit cell a 'plaquette' label p. Hence each star and each plaquette is associated with a different set of four spins. Consider the Hamiltonian

$$\hat{H} = -J_e \sum_{\text{stars}} \hat{A}_s - J_m \sum_{\text{plaquettes}} \hat{B}_p \tag{112}$$

where

$$\hat{A}_s = \prod_{j \in s} \hat{\sigma}_j^x; \qquad \hat{B}_p = \prod \hat{\sigma}_j^z. \tag{113}$$

Note that all the parts commute:

$$[\hat{A}_s, \hat{A}_{s'}] = [\hat{B}_p, \hat{B}_{p'}] = [\hat{A}_s, \hat{B}_p] = 0 \quad \forall s, s', p, p'. \tag{114}$$

The first two are obvious; the third comes about because a star and a plaquette share either zero or two common spins.

4.2.2 The ground state

Since the terms commute we can look for states which are eigenstates of all of them. The ground states of $\hat{B}_p$ are those possessing spin quantum numbers s_j (eigenvalues of $\hat{\sigma}_j^z$) that multiply to $+1$ around plaquette p: defining

$$w_l(s) = \prod_{j \in s} s_j \tag{115}$$

for any loop l, we have $w_p = +1$ for all plaquettes p. Hence the ground state is

$$|\Psi\rangle = \sum_{s: w_p = 1 \, \forall p} c_s |s\rangle. \tag{116}$$

(Notice that the state $|\Psi\rangle$ could also be produced by measuring the operators $\hat{A}_s$ and $\hat{B}_p$ and selecting the measurement result $+1$ in all cases.)

Eigenstates of $\hat{A}_s$ with eigenvalue $+1$ can be found if c_s is equal for each orbit of the action of $\hat{A}_s$ on a single site. If the system is a plane, this orbit is the entire plane and all orbits are equivalent (all can be shrunk to a point). If there are topologically distinct classes of orbit then the coefficients c are constrained to be equal for all $\{s\}$ such that the sums of s along each type of orbit is the same. So for a manifold of genus g (i.e. containing g 'holes' and $2g$ non-equivalent loops) the ground state has a degeneracy 4^g, corresponding to $2g$ qubits.

The fact that the toric code encodes two qubits can also be deduced from the general theory of quantum error correction (see problem sheet).

4.2.3 Stability of the states

Consider a general perturbing Hamiltonian with one- and two-qubit terms:

$$\hat{V} = -\sum_j \vec{h}_j \cdot \vec{\sigma}_j + \sum_{ij} J_{ij}(\vec{\sigma}_i, \vec{\sigma}_j). \tag{117}$$

The matrix elements of $\hat{V}$ are zero within the protected subspace up to order $k/2$, since repeated applications of $\hat{V}$ must produce a path spanning the whole torus to achieve a non-zero result. Hence the effects of the perturbation die off faster than any power of k as $k \to \infty$ (they decay as e^{-ak}).

4.2.4 The excitations

The elementary excitations of the model are of two types: we can violate star constraints ($A_s = -1$) or plaquette constraints ($B_p = -1$). Since

$$\prod_s \hat{A}_s = 1; \qquad \prod_p \hat{B}_p = 1 \tag{118}$$

(see problem sheet) it's impossible to violate just a single constraint. Instead, apply one of the 'path operators'

$$\hat{W}_c^{(e)} = \prod_{j \in c} \hat{\sigma}_j^z; \quad \hat{W}_{c'}^{(m)} = \prod_{j \in c'} \hat{\sigma}_j^x. \tag{119}$$

where c is a path on the lattice (i.e. between 'star' vertices) and c' is a path on the dual lattice (i.e. between plaquettes). A spin j is taken to be included in the path whenever it crosses the midpoint of a link. These commute respectively with all star operators except those at the end-point of the chain c, and with all plaquette operators except the end points of c'. We call these 'electric particles' (at stars) and 'magnetic particles' (at plaquettes). An electric particle pair has energy $2J_e$, and a magnetic pair has energy $2J_m$ (note the strictly non-zero gap from the ground state). In the pure toric code model the exceptions are perfectly localised (they have infinite effective mass); this effective mass will be made finite by a perturbation. The exponentially small effect of the local perturbation described above can be thought of as arising through the tunnelling of virtual excitations.

Now we study the statistics of these excitations. Imagine keeping a magnetic particle fixed at point p and 'braiding' an electric particle around it (i.e. choosing a curve c enclosing p). This corresponds to acting on the state $|\xi\rangle$ containing the magnetic particle as

$$|\xi\rangle \to \prod_{j \in c} \hat{\sigma}_j^{(z)} |\xi\rangle = \prod_{p' \,\text{within}\, c} \hat{B}'_p |\xi\rangle = -|\xi\rangle, \tag{120}$$

a minus sign being produced since $\hat{B}_p|\xi\rangle = -|\xi\rangle$. We say they have a nontrivial 'mutual statistics'.

Particles of the same type commute (i.e. are bosons): not only do they braid around each other to give +1 but they also exchange to give a sign +1 (an operation that is not properly defined for non-identical particles). Hence (see diagram) exchange of bound *em* pairs produces a minus sign, and these pairs are fermions.

Since pairs of electric and magnetic particles can be created and destroyed by local operations there are four particle sectors ('superselection sectors') that can be connected only by highly nonlocal operations: these are the vacuum sector 1, the electric sector e, the magnetic sector m, and the composite (fermonic) sector $\epsilon = e \times m$. Creation and destruction of particles gives 'fusion rules', for example $1 \times e = e$, $e \times e = 1$, $e \times m = \epsilon$ etc.

4.3 The need for non-abelian anyons

The excitations of the toric code acquire a non-trivial phase when they are 'braided' around one another, so they do not have either conventional Bose or Fermi statistics.

We can deduce the existence of a ground-state degeneracy, and of operators that can manipulate these states, from the statistics alone. Let $\hat{Z}$ create a pair of electric charges, propagate one of them around the torus, and annihilate them again; in the toric code this is just

$$\hat{Z} = \hat{W}^{(e)}_{c_1} \tag{121}$$

where c_1 is one of the nontrivial paths spanning the torus. Similarly let $\hat{X}$ create a magnetic pair and propagate them around an inequivalent spanning loop:

$$\hat{X} = \hat{W}^{(m)}_{c_2'} \tag{122}$$

Then the combination $Z^{-1}X^{-1}ZX$ just corresponds to braiding one particle type around the other, hence to the mutual statistics of -1:

$$Z^{-1}X^{-1}ZX = -1 \quad \Rightarrow \quad XZ = -ZX. \tag{123}$$

In other words there are non commuting operators Z and X for a single qubit—this at once tells us that the ground state is more than one-dimensional (i.e. is degenerate). Another non-commuting pair can be formed from the opposite choices of non-trivial paths (c_2 and c_1'). This gives us the ability to 'reach in' to the topologically protected space and manipulate its states by moving quasiparticles around. But unfortunately the algebra of the operators is not rich enough for universal quantum computation: we have $\{\hat{X}_A, \hat{Z}_A, \hat{X}_B, \hat{Z}_B\}$ but no two-qubit operator such as $\hat{Z}_A\hat{Z}_B$.

To get a universal gate set from natural excitations we require a state with *non-Abelian* anyonic excitations, where each braiding operation among the particles is associated with a nontrivial matrix operation on the manifold of ground states, not simply with multiplication by a phase. These matrix operations ('higher-dimensional representations of the braid group') will not commute with one another. Some such systems are rich enough to support universal quantum computation [42]. Candidates include

- The fractional quantum Hall state at $\nu = 5/2$ (the Moore-Read state);
- States of $p_x + \mathrm{i}p_y$ superconductors;
- Majorana fermions.

This is a highly active field, given the progress in superconducting qubits described above in §2.5.6 and the recent claimed observation of Majorana fermions in hybrid semiconductor-superconductor structures [43].

References

[1] Christopher M Dawson and Michael A Nielsen. The solovay-kitaev algorithm. *arXiv preprint quant-ph/0505030*, 2005.

[2] Y Nakamura and YA Pashkin. Coherent control of macroscopic quantum states in a single-Cooper-pair box. *Nature*, 1999.

[3] D Vion, A Aassime, A Cottet, P Joyez, and H Pothier. Manipulating the quantum state of an electrical circuit. *Science*, 2002.

[4] Matthias Steffen, M Ansmann, R McDermott, N Katz, Radoslaw Bialczak, Erik Lucero, Matthew Neeley, E Weig, A Cleland, and John Martinis. State Tomography of Capacitively Shunted Phase Qubits with High Fidelity. *Physical Review Letters*, 97(5), August 2006.

[5] J H Plantenberg, P C de Groot, C J P M Harmans, and J E Mooij. Demonstration of controlled-NOT quantum gates on a pair of superconducting quantum bits. *Nature*, 447(7146):836–839, June 2007.

[6] J Majer, J M Chow, J M Gambetta, Jens Koch, B R Johnson, J A Schreier, L Frunzio, D I Schuster, A A Houck, A Wallraff, A Blais, M H Devoret, S M Girvin, and R J Schoelkopf. Coupling superconducting qubits via a cavity bus. *Nature*, 449(7161):443–447, September 2007.

[7] M. D. Reed, L DiCarlo, S E Nigg, L Sun, L Frunzio, S M Girvin, and R J Schoelkopf. Realization of three-qubit quantum error correction with superconducting circuits. *Nature*, 482(7385):382–385, February 2012.

[8] Matteo Mariantoni, H Wang, T Yamamoto, M Neeley, Radoslaw C Bialczak, Y Chen, M Lenander, Erik Lucero, A D O'Connell, D Sank, M Weides, J Wenner, Y Yin, J Zhao, A N Korotkov, A N Cleland, and John M Martinis. Implementing the Quantum von Neumann Architecture with Superconducting Circuits. *Science*, 2011.

[9] Erik Lucero, Rami Barends, Yu Chen, Julian Kelly, Matteo Mariantoni, Anthony Megrant, Peter O'Malley, Daniel Sank, Amit Vainsencher, James Wenner, Ted White, Yi Yin, Andrew N Cleland, and John M Martinis. Computing prime factors with a Josephson phase qubit quantum processor. *arXiv.org*, quant-ph, February 2012.

[10] R Barends, J Kelly, A Megrant, D Sank, E Jeffrey, Y Chen, Y Yin, B Chiaro, J Mutus, C Neill, P O'Malley, P. Roushan, J Wenner, T C White, A N Cleland, and John M Martinis. Coherent Josephson qubit suitable for scalable quantum integrated circuits. *arXiv.org*, April 2013.

[11] R Barends, J Kelly, A Megrant, A Veitia, D Sank, E Jeffrey, T C White, J Mutus, A G Fowler, and B Campbell. Logic gates at the surface code threshold: Superconducting qubits poised for fault-tolerant quantum computing. *arXiv preprint arXiv:1402.4848*, 2014.

[12] Gavin W Morley, Marc Warner, A Marshall Stoneham, P Thornton Greenland, Johan van Tol, Christopher W M Kay, and Gabriel Aeppli. The initialization and manipulation of quantum information stored in silicon by bismuth dopants. *Nature Materials*, 9(9):725–729, January 2010.

[13] Alexei M Tyryshkin. Electron spin coherence exceeding seconds in high-purity silicon. *Nature Materials*, 11(2):143–147, December 2011.

[14] M Steger, K Saeedi, M L W Thewalt, J. J. L. Morton, H Riemann, N V Abrosimov, P Becker, and H J Pohl. Quantum Information Storage for over 180 s Using Donor Spins in a 28Si "Semiconductor Vacuum". *Science*, 336(6086):1280–1283, June 2012.

[15] Andrea Morello, Jarryd J Pla, Floris A Zwanenburg, Kok W Chan, Kuan Y Tan, Hans Huebl, Mikko Mottonen, Christopher D Nugroho, Changyi Yang, Jessica A van Donkelaar, Andrew D C Alves, David N Jamieson, Christopher C Escott, Lloyd C L Hollenberg, Robert G Clark, and Andrew S Dzurak. Single-shot readout of an electron spin in silicon. *Nature*, 467(7316):687–691, January 2010.

[16] Stephanie Simmons, Richard M Brown, Helge Riemann, Nikolai V Abrosimov, Peter Becker, Hans-Joachim Pohl, Mike L W Thewalt, Kohei M Itoh, and John J L Morton. Entanglement in a solid-state spin ensemble. *Nature*, 469(7332):69–72, January 2011.

[17] Jarryd J Pla, Kuan Y Tan, Juan P Dehollain, Wee H Lim, John J L Morton, Floris A Zwanenburg, David N Jamieson, Andrew S Dzurak, and Andrea Morello. High-fidelity readout and control of a nuclear spin qubit in silicon. *Nature*, 496(7445):334–338, April 2013.

[18] F Jelezko, T Gaebel, I Popa, A Gruber, and J Wrachtrup. Observation of Coherent Oscillations in a Single Electron Spin. *Physical Review Letters*, 92(7):076401, February 2004.

[19] M V G Dutt, L Childress, L Jiang, E Togan, J Maze, F Jelezko, A S Zibrov, P R Hemmer, and M D Lukin. Quantum Register Based on Individual Electronic and Nuclear Spin Qubits in Diamond. *Science*, 316(5829):1312–1316, June 2007.

[20] G Waldherr, Y. Wang, S Zaiser, M Jamali, T Schulte-Herbrüggen, H Abe, T Ohshima, J Isoya, J F Du, P Neumann, and J Wrachtrup. Quantum error correction in a solid-state hybrid spin register. *Nature*, 506(7487):204–207, February 2014.

[21] A. M. Stoneham, A J Fisher, and P T Greenland. Optically driven silicon-based quantum gates with potential for high-temperature operation. *Journal Of Physics-Condensed Matter*, 15:L447, 2003.

[22] P T Greenland, S A Lynch, A F G van der Meer, B N Murdin, C R Pidgeon, B Redlich, N Q Vinh, and G Aeppli. Coherent control of Rydberg states in silicon. *Nature*, 465(7301):1057–1061, March 2010.

[23] A C Johnson, J R Petta, J M Taylor, A Yacoby, M D Lukin, C M Marcus, M P Hanson, and A C Gossard. Triplet–singlet spin relaxation via nuclei in a double quantum dot. *Nature*, 435(7044):925–928, June 2005.

[24] F H L Koppens, C Buizert, K J Tielrooij, I T Vink, K C Nowack, T Meunier, L P Kouwenhoven, and L M K Vandersypen. Driven coherent oscillations of a single electron spin in a quantum dot. *arXiv.org*, cond-mat.mes-hall, August 2006.

[25] D J Reilly, J M Taylor, J R Petta, C M Marcus, M P Hanson, and A C Gossard. Suppressing Spin Qubit Dephasing by Nuclear State Preparation. *Science*, 321(5890):817–821, August 2008.

[26] A N Vamivakas, C Y Lu, C Matthiesen, Y Zhao, S Falt, A Badolato, and M Atatüre. Observation of spin-dependent quantum jumps via quantum dot resonance fluorescence. *Nature*, 467(7313):297–300, September 2010.

[27] J Brooke, D Bitko, TF Rosenbaum, and G Aeppli. Quantum annealing of a disordered magnet. *Science*, 284(5415):779–781, January 1999.

[28] GE Santoro, R Martonak, E Tosatti, and R Car. Theory of quantum annealing of an Ising spin glass. *Science*, 295(5564):2427–2430, 2002.

[29] GE Santoro and E Tosatti. Optimization using quantum mechanics: quantum annealing through adiabatic evolution. *Journal of Physics A: Mathematical and ...*, 2006.

[30] Ben W Reichardt. The quantum adiabatic optimization algorithm and local minima. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing - STOC '04*, pages 502–510, New York, New York, USA, 2004. ACM Press.

[31] E Farhi, J Goldstone, S Gutmann, J Lapan, A Lundgren, and D Preda. A quantum adiabatic evolution algorithm applied to random instances of an NP-complete problem. *Science*, 292(5516):472, 2001.

[32] W. van Dam, M. Mosca, and U. Vazirani. How powerful is adiabatic quantum computation, 2002. *Arxiv preprint quant-ph/0206003*, 2002.

[33] Dorit Aharonov, Wim van Dam, Julia Kempe, Zeph Landau, Seth Lloyd, and Oded Regev. Adiabatic Quantum Computation is Equivalent to Standard Quantum Computation. *arXiv.org*, quant-ph, May 2004.

[34] Julia Kempe, Alexei Kitaev, and Oded Regev. The Complexity of the Local Hamiltonian Problem. *arXiv.org*, quant-ph, June 2004.

[35] E Bernstein and U. Vazirani. Quantum complexity theory. *Siam Journal on Computing*, 26(5):1411–1473, 1997.

[36] Boris Altshuler, Hari Krovi, and Jeremie Roland. Adiabatic quantum optimization fails for random instances of NP-complete problems. *arXiv.org*, quant-ph, August 2009.

[37] Neil G Dickson and M. H. S. Amin. Does Adiabatic Quantum Optimization Fail for NP-Complete Problems? *Physical Review Letters*, 106(5):–, 2011.

[38] M Amin, C Truncik, and D Averin. Role of single-qubit decoherence time in adiabatic quantum computation. *Physical Review A*, 80(2), August 2009.

[39] M. W. Johnson, M. H. S. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. J. Berkley, J. Johansson, P. Bunyk, E. M. Chapple, C. Enderud, J. P. Hilton, K. Karimi, E. Ladizinsky, N. Ladizinsky, T. Oh, I. Perminov, C. Rich, M. C. Thom, E. Tolkacheva, C. J. S. Truncik, S Uchaikin, J. Wang, B. Wilson, and G. Rose. Quantum annealing with manufactured spins. *Nature*, 473(7346):194–198, 2011.

[40] A Yu Kitaev. Fault-tolerant quantum computation by anyons. *Annals Of Physics*, 303(1):2–30, January 2003.

[41] A G Fowler, M Mariantoni, J M Martinis, and A N Cleland. Surface codes: Towards practical large-scale quantum computation. *Physical Review A*, 86(032324), 2012.

[42] Chetan Nayak, Steven Simon, Ady Stern, Michael Freedman, and Sankar das Sarma. Non-Abelian anyons and topological quantum computation. *Reviews Of Modern Physics*, 80(3):1083–1159, September 2008.

[43] V Mourik, K Zuo, S M Frolov, S R Plissard, E P A M Bakkers, and L P Kouwenhoven. Signatures of Majorana Fermions in Hybrid Superconductor-Semiconductor Nanowire Devices. *Science*, 336(6084):1003–1007, May 2012.

Physics by the Lake 2015
Problems for QIP Lectures

Definitions of the Pauli matrices:

$$\sigma_x = \hat{X} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \qquad \sigma_y = \hat{Y} = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}; \qquad \sigma_z = \hat{Z} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$\sigma_+ = \frac{1}{2}(\sigma_x + \mathrm{i}\sigma_y) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}; \qquad \sigma_- = \frac{1}{2}(\sigma_x - \mathrm{i}\sigma_y) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

1. **The Bloch sphere** Show that the most general density operator for a single spin-1/2 particle can be written as

$$\hat{\rho} = \frac{1}{2}\left[\hat{I} + \vec{\alpha} \cdot \vec{\sigma}\right],$$

where $\vec{\alpha}$ is a vector in ordinary three-dimensional space and $\vec{\sigma}$ is the vector of the three Pauli operators.

What are the expectation values of the three components of the spin? What condition must α satisfy in order for $\hat{\rho}$ to be a (a) a pure state; (b) a physically acceptable mixed state?

[Hint—you can use the freedom in your choice of axes to make the calculation easier.]

2. **Decoherence of a single qubit: the depolarising channel.** Consider a quantum channel acting on a single qubit corresponding to an operation $\mathcal{E}$ with the Kraus operators

$$\hat{E}_0 = \sqrt{1-p}\hat{I}; \quad \hat{E}_1 = \sqrt{\frac{p}{3}}\hat{X}; \quad \hat{E}_2 = \sqrt{\frac{p}{3}}\hat{Y}; \quad \hat{E}_3 = \sqrt{\frac{p}{3}}\hat{Z}. \tag{1}$$

Verify that the Kraus operators do indeed obey the condition

$$\sum_k \hat{E}_k^\dagger \hat{E}_k = \hat{I}. \tag{2}$$

Show also that for an arbitrary density operator $\hat{\rho}$ the result of $\mathcal{E}$ can be written

$$\mathcal{E}(\hat{\rho}) = \frac{p'}{2}\hat{I} + (1-p')\hat{\rho}, \tag{3}$$

i.e. it corresponds to throwing away the density operator with probability p' and replacing it with a completely random mixed state $\hat{I}/2$, otherwise doing nothing. How is p' related to p?

What is the effect of this operation on the Bloch sphere introduced in Q1?

3. **A simple repetition code.** Consider the 5-qubit repetition code in which the logical states are encoded in states of five physical qubits:

$$|0_L\rangle = |00000\rangle; \qquad |1_L\rangle = |11111\rangle.$$

Write down the number of independent stabiliser generators and identify a suitable set of generators. How many bit-flip errors can it detect? How many phase errors? What is the distance of the code?

4. **Deviations from the adiabatic theorem.** Consider a system with a Hamiltonian $\hat{H}(s)$ dependent on a parameter s, and a set of instantaneous orthonormal eigenstates $|\psi_n(s)\rangle$ and eigenvalues $E_n(s)$. Show that

$$\langle\psi_m(s)|\frac{\mathrm{d}\psi_n(s)}{\mathrm{d}s}\rangle = \frac{\langle\psi_m(s)|\frac{\mathrm{d}\hat{H}(s)}{\mathrm{d}s}|\psi_n\rangle}{E_m(s)-E_n(s)}.$$

Hence show that if the eigenstates $|\psi_n(s)\rangle$ are used as a basis to expand a general state of the system in the form

$$\hat{\rho} = \sum_{mn}\rho_{mn}(t)|\psi_m(s(t))\rangle\langle\psi_n(s(t))|\mathrm{e}^{\mathrm{i}(E_n-E_m)t},$$

(taking $\hbar = 1$), then the expansion coefficients ρ_{mn} satisfy

$$\frac{\mathrm{d}\rho_{mn}}{\mathrm{d}t} = -\frac{\mathrm{d}s}{\mathrm{d}t}\sum_k\left[\frac{\langle\psi_m(s)|\frac{\mathrm{d}\hat{H}(s)}{\mathrm{d}s}|\psi_k\rangle}{E_m(s)-E_k(s)}\rho_{kn}\mathrm{e}^{\mathrm{i}(E_m-E_k)t} + \rho_{mk}\frac{\langle\psi_k(s)|\frac{\mathrm{d}\hat{H}(s)}{\mathrm{d}s}|\psi_n\rangle}{E_n(s)-E_k(s)}\mathrm{e}^{\mathrm{i}(E_n-E_k)t}\right].$$

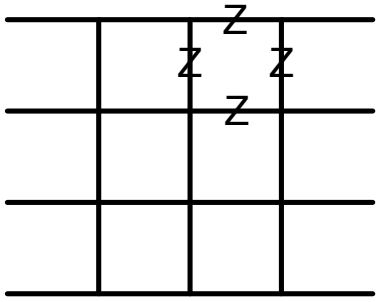
5. **Stabilisers for the toric code.** Show that all the star and plaquette operators $\hat{A}_s$ and $\hat{B}_p$ defined in the lectures are individually stabilisers for the ground-state manifold of the toric code. Show also that

$$\prod_s \hat{A}_s = 1; \qquad \prod_p \hat{B}_p = 1.$$

Hence show that on a $l\times l$ square lattice with periodic boundary conditions (a torus) there are $2(l^2-1)$ independent stabilisers.

By using the general relationships stated in the lectures between the number of stabilisers, the size n of the code and the number k of encoded qubits, deduce that the toric code encodes exactly two qubits.

6. **Surface codes.** A planar surface code is similar to a toric code but has open (rather than periodic) boundary conditions. It is defined on a $n\times(n-1)$ planar lattice with qubits assigned as usual to edges on the lattice. The lattice has "rough" and "smooth" edges as indicated in the image below, which also shows one of the plaquette operators.

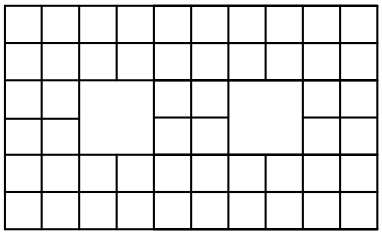


(a) On an $n\times(n-1)$ lattice, with rough and smooth edges as indicated in the figure, how many independent vertex operators are there? How many independent plaquette operators are there? Thus derive the number of encoded qubits in this code.

(b) Draw a sketch indicating logical X and Z operators for an encoded qubit.

(c) What is the distance of a planar surface code defined on a $n \times (n-1)$ lattice? Justify your answer.

(d) Planar surface codes can also be defined on lattices containing interior boundaries or holes. One such lattice is indicated in the figure below. Determine the number of encoded qubits in this code and identify logical X and Z operators for *one* of these qubits. *Hint:* There are 52 plaquettes, 75 vertices and 128 edges in this lattice.



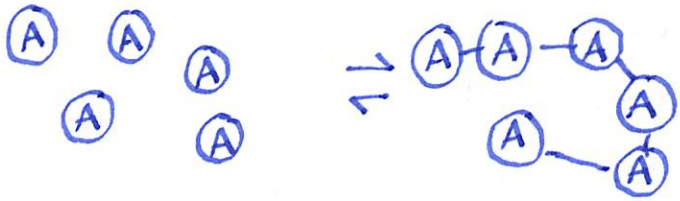
SCM

Physics of Soft Materials: Polymer Physics.

For detailed notes see online.

- Basic polymerisation process:

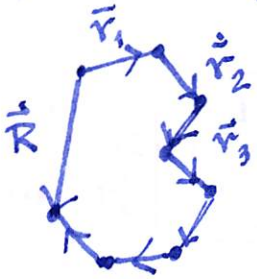
$$n\{A\} \rightleftharpoons (A)_n .$$



Connectivity introduces
- Long range effects.
- Strong fluctuations.
- high susceptibility to external force.
- Long term memory effects.

1. Static properties of polymers

1.1. Freely jointed chain model: (Ideal gas) [FJC]



- N links of length b_0 (points at any direction independently).
- Represented by (N+1) position vectors.

$$\{\vec{R}_n\} = \{\vec{R}_0, \vec{R}_1 \ldots \vec{R}_N\}.$$

- Alternate equivalent description (bond vectors)

$$\{\vec{r}_n\} = \vec{R}_n - \vec{R}_{n-1} \quad ; \quad n = 1, 2 \ldots N$$

- Probability distribution $\Rightarrow \Psi(\{\vec{r}_n\}) = \prod_{n=1}^{N} \Psi(\vec{r}_n)$
(Since bond vectors are independent of each other)

$\Psi(\vec{r})$: random distribution of vector of constant length b_0.

$$\boxed{\Psi(\vec{r}) = \frac{1}{4\pi b_0^2}\,\delta(|\vec{r}| - b_0)} \quad ; \quad \boxed{\int \Psi(r)\, dr = 1.}$$

- Important measure: End to end distance of the polymer. End to end vector $\vec{R} = \vec{R}_N - \vec{R}_0 = \sum_{n=1}^{N} r_n$

define $\vec{R} = \langle \vec{R}^2 \rangle^{1/2} = \langle (\vec{R}_N - \vec{R}_0)^2 \rangle^{1/2}$

$$\langle \vec{R}^2 \rangle = \sum_{n,m=1}^{N} \langle \vec{r}_n \cdot \vec{r}_m \rangle = \sum_{n=1}^{N} \langle \vec{r}_n^{\,2} \rangle + 2 \sum_{n>m} \langle \vec{r}_n \cdot \vec{r}_m \rangle$$

$$\langle \vec{R}^2 \rangle = N b_0^2 \quad \text{since} \quad \langle \vec{r}_n \cdot \vec{r}_m \rangle = \langle \vec{r}_n \rangle \cdot \langle \vec{r}_m \rangle = 0 .$$

Note $\bar{R} = \sqrt{N}\, b_0$.

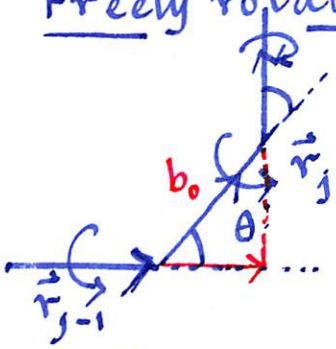
$\boxed{\bar{R} \sim N^{1/2}}$

N: measure of molecular weight

Note: FJC: off lattice model. Random walk on square lattice gives exactly same answer (no self-avoidance)

This scaling is true even for general models e.g.

Freely rotating chain. (FRC).



- n^{th} bond connected to $(n-1)^{th}$ bond at fixed angle θ.
- ϕ: azimuthal angle random.

$$\langle \vec{r}_n \cdot \vec{r}_m \rangle = \cos\theta \langle \vec{r}_{n-1} \cdot \vec{r}_m \rangle = \cos^2\theta \langle \vec{r}_{n-2} \cdot \vec{r}_m \rangle$$

and so on. and so forth.

Thus $\langle \vec{r}_n \cdot \vec{r}_m \rangle = b_0^2 (\cos\theta)^{|n-m|}$ $\left[\because \langle \vec{r}_m^{\,2} \rangle = b_0^2\right]$

$$\boxed{\langle \vec{R}^2 \rangle = \sum_{n=1}^{N} \sum_{m=1}^{N} \langle \vec{r}_n \cdot \vec{r}_m \rangle = N b_0^2 \left[\frac{1+\cos\theta}{1-\cos\theta}\right]}$$

Same scaling form.

Rewrite $\langle \vec{r}_n \cdot \vec{r}_m \rangle = b_0 (\cos\theta)^{|n-m|}$

$(\cos\theta)^{|n-m|} = \exp\left[|j-i| \ln\cos\theta\right] = e^{-\left\{\frac{|j-i|}{s_p}\right\}}$

where $s_p = -1/\ln(\cos\theta)$ result shows correlations decay strongly with $|j-i|$.

Worm like chain model: (Kratky Porod model).

$$\{\Psi(\vec{r}_n)\} = \prod_{n=1}^{N} \Psi(\vec{r}_n) \; ; \text{ where } \Psi(\vec{r}_n) = \frac{1}{4\pi b_0^2} \delta(|\vec{r}_n| - b) \times e^{-\frac{E\{\vec{r}_n\}}{k_B T}}$$

$$E\{\vec{r}_n\} = -\kappa \sum_{n=1}^{N} \hat{t}_n \cdot \hat{t}_{n-1} \; ; \text{ where } \hat{t}_n = \frac{\vec{r}_n}{|\vec{r}_n|} : \text{unit tangent vector.}$$

The problem is equivalent to 1-d Heisenberg model; Partition sum can be evaluated exactly.

$$Z_{KPC} = \int \prod_{i=1}^{N-1} d\cos\theta_i \exp\left[\frac{\kappa}{k_B T} \sum_{i=1}^{N} \cos\theta_i\right] = \left(\frac{2}{\tilde{\kappa}} \sinh\tilde{\kappa}\right)^N$$

$\tilde{\kappa} = \kappa / k_B T$.

One can use the Kratky Porod partition function to evaluate the correlation function $\langle \hat{t}_n \cdot \hat{t}_m \rangle$ and thereby $\langle \vec{R}^2 \rangle$

$$\langle \hat{t}_n \cdot \hat{t}_m \rangle = b_0^2 \left[\coth \tilde{\kappa} - \frac{1}{\tilde{\kappa}} \right]^{|n-m|} \equiv b_0^2 \exp\left[-\frac{b_0}{\xi} |n-m| \right]$$

$\xi \rightarrow$ persistence length of the chain. depends on the chain rigidity.

$$\xi = \frac{b_0}{\ln\left[\coth \kappa - \frac{1}{\kappa} \right]} \approx b_0 \tilde{\kappa} \quad \text{in the limit } \tilde{\kappa} \to \infty.$$

Continuum limit: In the continuum limit.

$$b_0 \sum_{n=1}^{N} \rightarrow \int_0^L du \quad \text{and} \quad b_0 \sum_{m=1}^{N} \rightarrow \int_0^L dv. \; ; \; L = \text{contour length.}$$

$$\langle \vec{R}^2 \rangle = b_0^2 \sum_{n=1}^{N} \sum_{m=1}^{N} \exp\left[-\frac{b_0}{\xi} |n-m| \right]$$

$$\boxed{\langle \vec{R}^2 \rangle = 2\xi L - 2\xi^2 \left(1 - \exp\left[-L/\xi \right] \right).}$$

(i) Ideal chain limit $\langle \vec{R}^2 \rangle \cong 2\xi L \quad L \gg \xi$.

(ii) Rod-limit expand $e^{-L/\xi} \cong 1 - \frac{L}{\xi} + \frac{1}{2}\left(\frac{L}{\xi}\right)^2 + \cdots$

Thus $\langle \vec{R}^2 \rangle \cong L^2 \quad L \ll \xi$.

$\Rightarrow \bar{R} = \langle \vec{R}^2 \rangle^{1/2} \cong L. \; (\propto N)$. { as opposed to $\sqrt{N}$ for ideal chains }

The Kratky-Porod model and its continuum version the wormlike chain has been successfully applied to understand equilibrium properties of biopolymers e.g. DNA.

Continuum models: $\Phi(\vec{R}, N)$: Probability distribution that end to end vector of N links is $\vec{R}$.

$$\Phi(\vec{R}, N) = \int d\vec{r}_1 \int d\vec{r}_2 \cdots \int d\vec{r}_N \; \delta\left(R - \sum_{n=1}^{N} \vec{r}_n \right) \Psi(\{\vec{r}_n\})$$

rewriting $\delta(\vec{r}) = \frac{1}{(2\pi)^3} \int d\vec{k} \, e^{i\vec{k}\cdot\vec{r}}$ { using this. $\Phi(\vec{R}, N)$ can be explicitly evaluated.

$$\Phi(\vec{R},N) = \frac{1}{(2\pi)^3}\int d\vec{k}\, e^{i\vec{k}\cdot\vec{R}} \left(\frac{\sin kb}{kb}\right)^N$$

(i) Large N $\left(\frac{\sin kb}{kb}\right)$ very small unless kb is small

(ii) for $kb \ll 1$ $\left(\frac{\sin kb}{kb}\right)^N \cong \left(1 - \frac{k^2b^2}{6}\right)^N \cong \exp\left[-\frac{Nk^2b^2}{6}\right]$

$$\Phi(\vec{R},N) \cong \left(\frac{3}{2\pi N b^2}\right)^{3/2} \exp\left[-\frac{3\vec{R}^2}{2Nb^2}\right]$$

→ End to end vector distribution is Gaussian.

One can also prove that if bond length distributions are Gaussian

i.e. $\Psi(\vec{r}) = \left[\frac{3}{2\pi b^2}\right]^{3/2} \exp\left[-\frac{3r^2}{2b^2}\right]$

then the distribution of vector $\vec{R}_n - \vec{R}_m$ between any two segments is also Gaussian.

$$\Phi(\vec{R}_n - \vec{R}_m, n-m) = \left[\frac{3}{2\pi b^2 |n-m|}\right]^{3/2} \exp\left[-\frac{3}{2}\frac{(\vec{R}_n - \vec{R}_m)^2}{|n-m| b^2}\right]$$

Radius of Gyration of a polymer: Also characterizes geometric measure of an isolated polymer. It is particularly useful to describe branched polymers.

$$R_G^2 = \frac{1}{N}\sum_{i=1}^{N}(\vec{R}_i - \vec{R}_{cm})^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\vec{R}_i - \frac{1}{N}\sum_{j=1}^{N}\vec{R}_j\right)^2$$

$$R_G^2 = \frac{1}{2N^2}\sum_{n,m} \langle (\vec{R}_n - \vec{R}_m)^2 \rangle$$ and in the continuum limit

$$\langle R_G^2 \rangle = \frac{1}{N^2}\int_0^N \int_0^N \langle (\vec{R}(u) - \vec{R}(v))^2 \rangle \, dv\, du$$

For an ideal linear chain $\langle (\vec{R}(u) - \vec{R}(v))^2 \rangle = (v-u)b^2$

$$\langle R_G^2 \rangle = Nb^2/6 = \frac{\langle \vec{R}^2 \rangle}{6}$$

→ connection between radius of gyration and end to end distance

Free energy of an ideal chain:

chain entropy $\Rightarrow S = k_B \ln \Phi(\vec{R}, N)$.

$$S = -\frac{3}{2} k_B \vec{R}^2 / N b^2 + \frac{3}{2} k_B \ln\left[\frac{3}{2\pi N b^2}\right]$$; note second term does not depend on $\vec{R}$

$$S(N, \vec{R}) = -\frac{3}{2} k_B \frac{\vec{R}^2}{N b^2} + S(N, 0)$$

$$F(N, \vec{R}) = U - TS = \frac{3}{2} k_B T \frac{\vec{R}^2}{N b^2}$$

↑ not important in soft systems.

II Equilibrium behaviour of real chains:

Our discussions on ideal chains neglect the important effect of long range interactions mediated by the chain backbone. Two monomers cannot occupy the same space. This causes excluded volume interactions $\Rightarrow$ monomer repulsion $\Rightarrow$ chain swelling.

CONSEQUENCES. $\langle \vec{R}^2 \rangle \sim \langle \vec{R}_G^2 \rangle \sim N^{2\nu}$ $\nu > \frac{1}{2}$

Thermodynamics of polymer chains in a solvent determined by interactions between monomers and solvent molecules. Ⓐ-Ⓐ interaction less than Ⓐ - Ⓢ interaction [where Ⓢ ≡ solvent] → chain swelling.

Flory theory of Polymers in good solvent.

- N monomers swollen to size $R > R_0 = bN^{1/2}$.
- Assume monomers uniformly distributed in volume R^3 with no correlations among them.
- Probability second monomer within excluded volume v of a given monomer $= v(N/R^3)$
- Energy $\sim k_B T$ per exclusion $\Rightarrow E \equiv k_B T v (N/R^3)$
- For N monomers $F_{int} \approx (N^2/R^3) v k_B T$
- Entropic contribution $F_{ent} \approx k_B T \frac{R^2}{N b^2}$.
- $F = F_{ent} + F_{int} = v k_B T \frac{N^2}{R^3} + k_B T R^2 / N b^2$

Minimum gives optimum size of real chain in Flory-theory ($\partial F/\partial R = 0$) $F = k_B T \left(\nu \frac{N^2}{R^3} + \frac{R^2}{N b^2} \right)$

$$\frac{\partial F}{\partial R} = 0 \Rightarrow k_B T \left(-3\nu \frac{N^2}{R_F^4} + 2 \frac{R_F}{N b^2} \right) \Rightarrow R_F^5 \approx \nu b^2 N^3$$

$R_F \approx \nu^{1/5} b^{2/5} N^{3/5}$ more accurate expressions of the exponent can be obtained employing heavy machinery of field theory. & extensive computer simulations Flory's estimate is surprisingly good approximation.

Flory exponent $\nu_F = 0.6$. $\nu_{MC} = 0.588\ldots$

Thermodynamics of Polymer solutions:

Species A : volume V_A ; Species B : volume V_B.

Volume fractions. $\phi_A = \frac{V_A}{V_A + V_B}$; $\phi_B = \frac{V_B}{V_A + V_B} = 1 - \phi_A$

Assume lattice volume = v_0.

Thus molecular volume of A = $V_A = N_A v_0$

Similarly B $V_B = N_B v_0$; Combined systems occupy n lattice sites. $n = \frac{V_A + V_B}{v_0}$ ← all molecules of A occupy V_A / v_0 lattice sites $= n \phi_A$.

Entropy $S = k_B \ln \Omega$: Ω : # of ways to arrange molecules on lattice

$\Omega_{AB} = n$; states for each molecule of a homogeneous A, B mixture.

$\Omega_A = n \phi_A$; For a single molecule mixing

$$\Delta S_A = k_B \ln \Omega_{AB} - k_B \ln \Omega_A = k_B \ln \left(\frac{\Omega_{AB}}{\Omega_A} \right) = -k_B \ln \Omega_A$$

$$\Delta S_{mix} = n_A \Delta S_A + n_B \Delta S_B = -k_B \left[n_A \ln \phi_A + n_B \ln \phi_B \right]$$

$n_A = \frac{n \phi_A}{N_A}$; $n_B = \frac{n \phi_B}{N_B}$; thus $\Delta \bar{S}_{mix} = \frac{\Delta S_{mix}}{n}$

$$\Delta \bar{S}_{mix} = -k_B \left[\frac{\phi_A}{N_A} \ln \phi_A + \frac{\phi_B}{N_B} \ln \phi_B \right].$$

$$\Delta \bar{S}_{mix} = -k_B \left[\frac{\phi_A}{N_A} \ln \phi_A + \phi_B \ln \phi_B \right]$$; for polymer solutions $N_A = N$ $N_B = 1$.

For polymer blends.

$$\Delta \bar{S}_{mix} = -k_B \left[\frac{\phi}{N_A} \ln \phi + \frac{(1-\phi)}{N_B} \ln (1-\phi) \right]$$

Energy of mixing:

Binary mixture $\phi_A = \phi$, $\phi_B = 1-\phi$

Ⓐ – Ⓐ interaction u_{AA}.
Ⓐ – Ⓑ interaction u_{AB}.
Ⓑ – Ⓑ interaction u_{BB}

$$U = \frac{zn}{2} \left[U_A \phi_A + U_B \phi_B \right]; \quad U_A = u_{AA} \phi_A + u_{AB} \phi_B.$$
$$U_B = u_{AB} \phi_A + u_{BB} \phi_B.$$

Noting the binary nature of mixture.

$$U = \frac{zn}{2} \left[u_{AA} \phi^2 + 2 u_{AB} (1-\phi) \phi + u_{BB} (1-\phi)^2 \right]$$

Total energy before mixing $\Rightarrow U_0 = \frac{zn}{2} \left[u_{AA} \phi + u_{BB} (1-\phi) \right]$

$$U - U_0 = \frac{zn}{2} \phi (1-\phi) (2 u_{AB} - u_{BB} - u_{AA})$$

$$\Delta \bar{U}_{mix} = \frac{U - U_0}{n} = \frac{z}{2} \phi (1-\phi) (2 u_{AB} - u_{AA} - u_{BB}) = \chi \phi (1-\phi) k_B T$$



χ : Flory interaction parameter.

$$\chi \equiv \frac{z}{2} \frac{(2 u_{AB} - u_{AA} - u_{BB})}{k_B T}.$$

$$\Delta \bar{F}_{mix} = \Delta \bar{U}_{mix} - T \Delta \bar{S}_{mix} = k_B T \left[\frac{\phi}{N_A} \ln \phi + \chi \phi (1-\phi) + \frac{(1-\phi)}{N_B} \ln (1-\phi) \right]$$

For polymer solutions $N_A = N$, $N_B = 1$. Note analogy with Bragg-Williams approximations in context of magnetic systems.

If $F_{mix}(\phi_0)$ lower than phase separated state energy $F_{\alpha\beta}(\phi_0)$ system remains mixed. Else the system undergoes phase separation with resulting component fractions following lever rule.

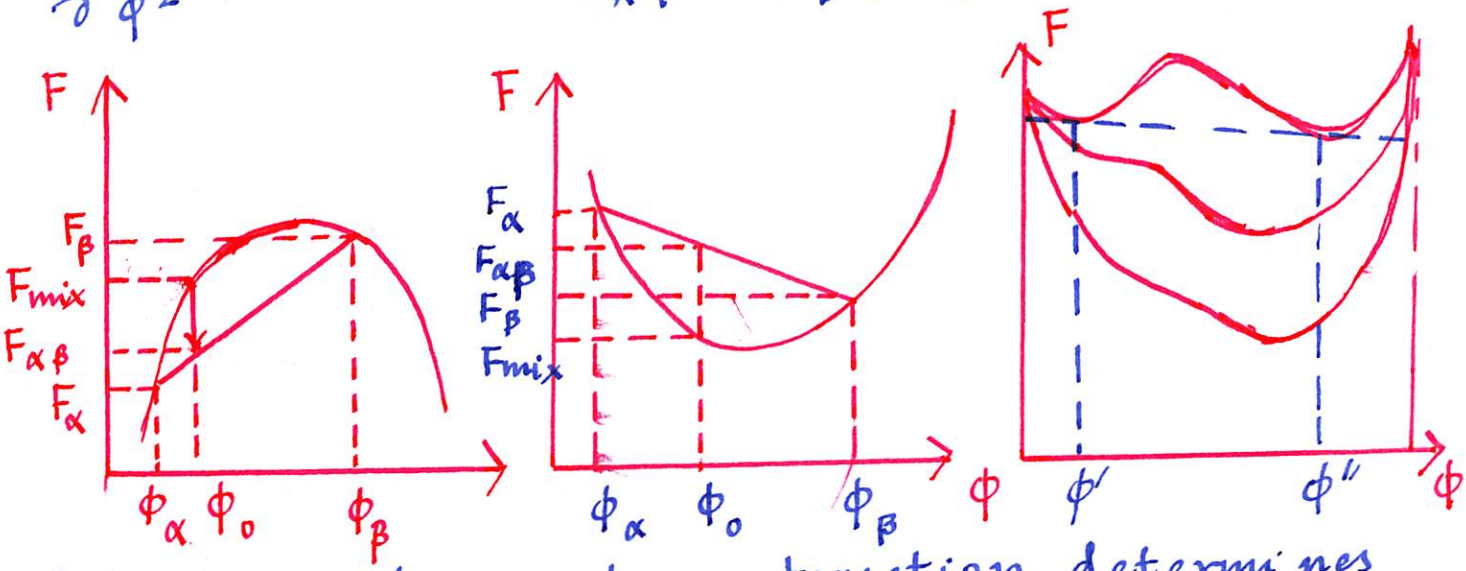
$\phi_0 = f_\alpha \phi_\alpha + f_\beta \phi_\beta$: note $f_\alpha = 1 - f_\beta$.

$f_\alpha = \dfrac{\phi_\beta - \phi_0}{\phi_\beta - \phi_\alpha}$; $f_\alpha = \dfrac{\phi_0 - \phi_\alpha}{\phi_\beta - \phi_\alpha}$; $F_{\alpha\beta}(\phi_0) = F_\alpha f_\alpha + F_\beta f_\beta$

Local stability criterion. $\dfrac{\partial^2 F_{mix}}{\partial \phi^2} < 0$ unstable

and $\dfrac{\partial^2 F}{\partial \phi^2} > 0 \rightarrow$ locally stable.

$$\frac{\partial^2 \Delta F_{mix}}{\partial \phi^2} = k_B T \left[\frac{1}{N_A \phi} + \frac{1}{N_B (1-\phi)} \right] - 2\chi k_B T$$



A common tangent construction determines compositions ~~in~~ of phase separated state.

For any overall composition in "miscibility gap" the system phase separates into ϕ & ϕ''. Within this gap there are unstable & metastable regions that are separated by line $(\partial^2 F/\partial \phi^2 = 0) \rightarrow$ spinodal.

Phase boundary determined by common tangent

$$\left.\frac{\partial \Delta F_{mix}}{\partial \phi}\right|_{\phi=\phi'} = \left.\frac{\partial \Delta F_{mix}}{\partial \phi}\right|_{\phi=\phi''} = k_B T \left[\ln \frac{\phi}{N} - \frac{\ln (1-\phi)}{N} + \chi (1-2\phi) \right] = 0$$

Binodal line.

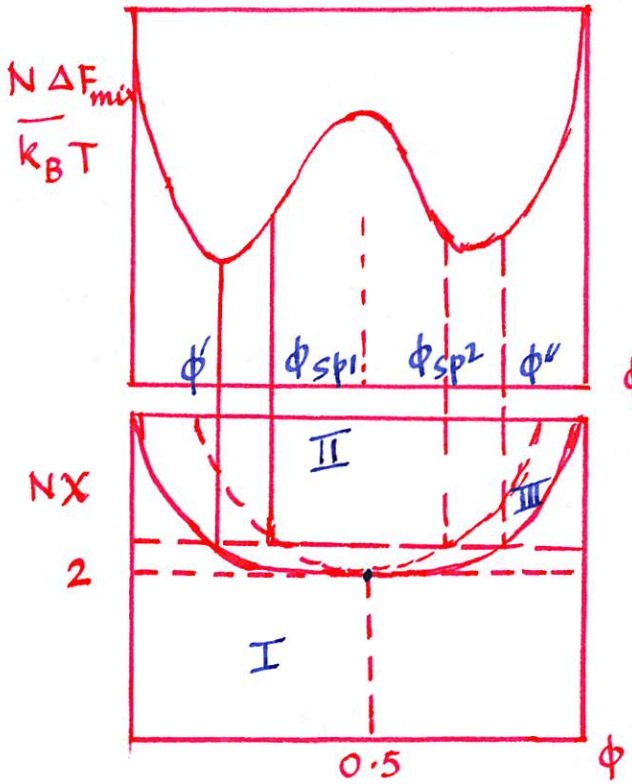
$$\frac{\partial^2 F_{min}}{\partial \phi^2} = 0 = k_B T \left[\frac{1}{N_A \phi} + \frac{1}{N_B (1-\phi)} - 2\chi \right] = 0 \rightarrow$$ Spinodal line.

Lowest point of spinodal = critical point

$$\chi_s = \frac{1}{2} \left[\frac{1}{N_A \phi} + \frac{1}{N_B (1-\phi)} \right] ; \quad \frac{\partial \chi_s}{\partial \phi} = 0 \Rightarrow + \frac{1}{2} \left[-\frac{1}{N_A \phi^2} + \frac{1}{N_B (1-\phi)^2} \right] = 0$$

$$\phi_c = \frac{\sqrt{N_B}}{\sqrt{N_A}+\sqrt{N_B}} \; ; \; \chi_c = \frac{1}{2}\frac{\left(\sqrt{N_A}+\sqrt{N_B}\right)^2}{N_A N_B} = \frac{1}{2}\left(\frac{1}{\sqrt{N_A}}+\frac{1}{\sqrt{N_B}}\right)^2$$

for symmetric polymer blends $\phi_c = \frac{1}{2}$ $\chi_c = \frac{2}{N}$



I : Single phase.

II : Two phase. (unstable)

III : Metastable

Bottom figure solid line : binodal

dashed line – spinodal.

Physics of Soft Matter : Membranes.

Amphiphilic molecules - ∩ → hydrophilic head. | - hydrophobic tail.



Monolayer membrane



Bilayer membrane

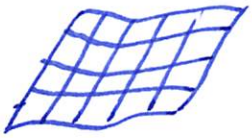
Such bilayers can form open as well as closed membranes (vesicles). The human red blood cell is one such example. An interesting phenomena observed in vesicles in particular in RBCs is a series of shape transformations. from



stomatocyte → discocyte → echinocyte.

The aim of this section would be to develop a theory of fluctuating interfaces.

Deformation modes of a membrane. (bilayer).



1) Stretch. 2) Bend. 3) Thickness. 4) Shear.

1. Stretch. $F_{Area} = \frac{K_a}{2 a_0} (\Delta a)^2$ — [Assume Hooke's law. a_0 - equilibrium area; Δa: change of area]

$K_a \cong$ Area stretch modulus $\cong 55 - 70\ k_B T / nm^2$

2. Bending : $F_{bend} = \frac{K_b}{2} \int ds\, (2H)^2 + \frac{K_G}{2} \int ds\, G$

H = Mean curvature $= \frac{1}{2}(c_1 + c_2)$ where c_1 & c_2 are principal curvatures.

G = Gaussian curvature $= c_1 c_2$.

3. Thickness change. $F_{thickness} = \frac{K_w}{2} \int ds \left(\frac{w - w_0}{w_0}\right)^2$

w : separation between bilayers.

w_0 : initial separation between leaflets.

4. Membrane shear : Fluid membranes of the type considered here do not support shear

For closed membranes the bending energy somewhat simplifies. Gauss-Bonnet theorem states

$$\int ds\, G = 4\pi (1-g).$$

G: Gaussian curvature

g: genus of surface.

$g = 0$ for a sphere.
$g = 1$ for torus.

This implies for a fixed topology Gaussian ~~curvature~~ curvature does not contribute to bending energy.

Thus $\boxed{F_{bend} = \frac{\kappa_b}{2} \int (2H)^2 ds.}$ ds: surface area element.

Spontaneous curvature model: $F_{bend} = \frac{\kappa_b}{2} \int (2H - C_0)^2 ds$

For this model the flat membrane is not the equilibrium shape.

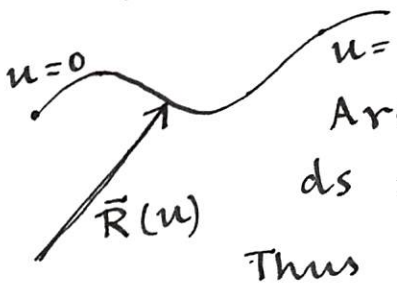
For shapes which have additional constraints e.g. if the total area is conserved one needs to add a suitable Lagrange multiplier to impose this.

Area difference elasticity model: Area difference between outer and inner leaflets $\Delta m = 2w \int ds\, H$.

$$F_{bend} = \frac{\kappa_b}{2} \int ds\, (2H - C_0)^2 + \frac{\kappa_b^* \pi}{8 A w^2} (\Delta m - \Delta m_0)^2$$

We now discuss the mathematical machinery with which to understand membrane geometry & mechanics.

Differential geometry of space curves and surfaces.



$\vec{r} = \vec{R}(u)$. u: parameter that denotes points along the curve

Arc length $\Delta s = |\vec{R}(u_1) - \vec{R}(u_2)|$

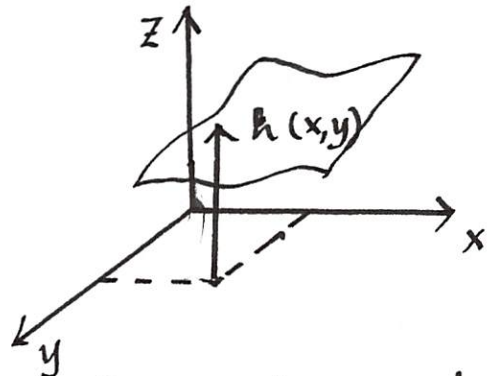
$ds = \left|\frac{\partial \vec{R}}{\partial u}\right| du$; tangent vector $\hat{t}(u) = \frac{d\vec{R}}{ds}$.

Thus $\hat{t}(u) = (\partial \vec{R}/\partial u)/\partial s/\partial u$

Rate of change of tangent related to curvature

$$\frac{d\hat{t}}{ds} = \kappa \hat{n}_c = \frac{d^2 \vec{R}}{ds^2}$$

Surfaces: Parametric form $x = f(u,v)$; $y = g(u,v)$
$z = h(u,v)$ which determine a vector $\vec{r}(u,v)$



Implicit form $F(x,y,z) = 0$.

Monge representation $u = x$, $v = y$
$z = h(x,y)$

$$\vec{r} \equiv (u, v, h(u,v)) = (x, y, h(x,y))$$

Tangent vectors $\vec{r}_u = d\vec{r}/du$
$\vec{r}_v = d\vec{r}/dv$ → define a tangent plane.

Normal vector $$\hat{n} = \frac{(\vec{r}_u \times \vec{r}_v)}{|\vec{r}_u \times \vec{r}_v|}$$

For an implicit form $F(x,y,z) = 0$. $$\hat{n} = \frac{\vec{\nabla} F}{|\vec{\nabla} F|}$$

Metric of a surface: Distance between points

$$ds^2 = (d\vec{r})^2 = (\vec{r}_u\, du + \vec{r}_v\, dv)^2$$
$$= E\, du^2 + 2F\, du\, dv + G\, dv^2$$

$E = \vec{r}_u^{\,2}$; $F = \vec{r}_u \cdot \vec{r}_v$; $G = \vec{r}_v^{\,2}$

Metric $g = EG - F^2 > 0$ since $(\vec{r}_u \times \vec{r}_v)^2 > 0$.

Area element $d^2s = dA = |\vec{r}_u \times \vec{r}_v|\, du\, dv$

Since $g = (\vec{r}_u \times \vec{r}_v)^2$ $dA = \sqrt{g}\, du\, dv$

Thus metric is used to measure lengths and areas

In Monge Gauge $\vec{r}_u = \vec{r}_x = (1, 0, h_x)$; $h_x \equiv \frac{\partial h(x,y)}{\partial x}$

Similarly $\vec{r}_v = \vec{r}_y = (0, 1, h_y)$. Thus area element

$$dA = dx\, dy \sqrt{1 + h_x^2 + h_y^2}$$

$$\hat{n} = \frac{\hat{z} - h_x \hat{x} - h_y \hat{y}}{\sqrt{1 + h_x^2 + h_y^2}}$$

Surface curvature: Normal curvature of a curve

$$\kappa = \vec{r}'' \cdot \hat{n}$$

$$\vec{r}'' = u''\vec{r}_u + v''\vec{r}_v + (u')^2 \vec{r}_{uu} + (v')^2 \vec{r}_{vv} + (2u'v')\vec{r}_{uv}$$

Note $\hat{n}\cdot\vec{r}_u = \hat{n}\cdot\vec{r}_v = 0$. thus curvature

$$\kappa = \frac{L\,du^2 + 2M\,du\,dv + N\,dv^2}{E\,du^2 + 2F\,du\,dv + G\,dv^2}$$

$L = \hat{n}\cdot\vec{r}_{uu}$ $\quad M = \hat{n}\cdot\vec{r}_{uv}$ $\quad N = \hat{n}\cdot\vec{r}_{vv}$

Also since $\hat{n}\cdot\vec{r}_u = \hat{n}\cdot\vec{r}_v = 0$ differentiating these expressions we get

$L = -\hat{n}_u\cdot\vec{r}_u$ $\quad M = -\hat{n}_v\cdot\vec{r}_u = -\hat{n}_u\cdot\vec{r}_v$; $N = -\hat{n}_v\cdot\vec{r}_v$

Thus $\kappa = -\dfrac{d\vec{r}\cdot d\hat{n}}{d\vec{r}\cdot d\vec{r}}$

Parametric representation : Mean & Gaussian curvature

$x = f(u,v)$, $y = g(u,v)$, $z = h(u,v)$

unit vector on surface $\hat{a} = \ell\,\vec{r}_u + m\,\vec{r}_v$

$\hat{a}^2 = 1 = E\ell^2 + 2F\ell m + Gm^2 = 1.$

$\therefore\ \kappa = L\ell^2 + 2M\ell m + Nm^2$. With constraint of unit vector maximise curvature.

$\tilde{\kappa} = L\ell^2 + 2M\ell m + Nm^2 - \lambda(E\ell^2 + 2F\ell m + Gm^2)$

$\partial\tilde{\kappa}/\partial\ell = 0 = \partial\tilde{\kappa}/\partial m$ Finding extremum values and plugging them in expression for κ. we have.

$\kappa^2(EG - F^2) - \kappa(EN + GL - 2FM) + (LN - M^2) = 0$

Roots κ_a and κ_b $\quad H = \frac{1}{2}(\kappa_a + \kappa_b)$; $K = \kappa_a\kappa_b$.

$$H = \frac{EN + GL - 2FM}{2(EG - F^2)} \quad ; \quad K = \frac{LN - M^2}{EG - F^2}$$

Principal directions: Two directions whose curvatures are extrema. One can eliminate Lagrange multiplier λ from $\partial\tilde{\kappa}/\partial\ell = 0$ $\partial\tilde{\kappa}/\partial m = 0$ to get

$(EM - FL)\ell^2 + (EN - GL)\ell m + (FN - GM)m^2 = 0.$

solutions for quantity (ℓ/m) such that $(\ell/m)_1 = \mu_1$ and $(\ell/m)_2 = \mu_2$ obeying

$$\mu_1 + \mu_2 = -\frac{EN - GL}{EM - FL} \quad ; \quad \mu_1\mu_2 = \frac{FN - GM}{EM - FL}.$$

Directions have unit vectors $\hat{a}_1 = m_1(\mu_1 \vec{r}_u + \vec{r}_v)$
$\hat{a}_2 = m_2(\mu_2 \vec{r}_u + \vec{r}_v)$

Equilibrium shapes: For a general functional $I = \int f[\psi(\vec{r}), \vec{\nabla}\psi(\vec{r})]$; the extrema conditions are found by variational calculus.

$$\frac{\delta I}{\delta \psi(\vec{r})} = 0 \Rightarrow \frac{\partial f}{\partial \psi(\vec{r})} - \vec{\nabla} \cdot \frac{\partial f}{\partial(\vec{\nabla}\psi(\vec{r}))} = 0.$$

For a "nearly" flat membrane. $F \cong \frac{\kappa}{2}\int d^2s\,(\vec{\nabla}^2 h)^2$

EL equations reduce to $\boxed{\kappa \nabla^4 h = 0}$. – Biharmonic eqn. pain to solve.

Fluctuations in fluid membranes.

In Monge Gauge $F_{bend} \cong \frac{1}{2}\kappa_{bend}(h_{xx}+h_{yy})^2$

In Fourier space $F_{bend} = \frac{1}{2A} k \sum_q q^4 |h_q|^2$ [$k = \kappa_{bend}$]

By equipartition theorem $\langle |h_q|^2 \rangle = \frac{T}{k q^4}$

Mean square fluctuation of height increases algaebrically with system size

$$\langle h(\vec{r})^2 \rangle = \frac{1}{A}\sum_{\vec{q}} \langle |h_q|^2 \rangle \sim \frac{T}{k} L^2$$

A quantity that can be probed in experiments is the normal-normal correlation function.

$$g_n(\vec{r}) = \langle (\hat{n}(\vec{r}) - \hat{n}(0))^2 \rangle \; ; \; \hat{n} \approx \hat{z} - h_x \hat{x} - h_y \hat{y}$$

$$g_n(\vec{r}) = \frac{2}{A}\sum_q q^2 \langle |h_q|^2 \rangle (1 - \cos \vec{q}\cdot\vec{r})$$

As $r \to \infty$ integral diverges logarithmically

$$\int_0^{2\pi} \cos(qr\cos\theta)\, d\theta = 2\pi J_0(qr) \; ; \; g_n(\vec{r}) \approx \frac{\alpha T}{4\pi k} \ln\left(\frac{r}{a}\right)$$

Physics of Soft Matter: Liquid Xtals.

Properties intermediate between liquid and crystal.
Liquid: inability to support shear, high fluidity etc.
crystal: anisotropic optical, magnetic, electrical properties.

Point in crystal $\vec{x} = \vec{x}_0 + n_1 \vec{a}_1 + n_2 \vec{a}_2 + n_3 \vec{a}_3$

Probability of finding a point stays finite as $|\vec{x} - \vec{x}_0| \to \infty$

Xtal $\lim_{|\vec{x} \to \vec{x}'| \to \infty} \langle \rho(\vec{x}) \rho(\vec{x}') \rangle = F(\vec{x} - \vec{x}') \to$ Periodic fn. of basis vector $\vec{a}_i$

Liquid $\lim_{|\vec{x} - \vec{x}'| \to \infty} \langle \rho(\vec{x}) \rho(\vec{x}') \rangle \simeq \bar{\rho}^2 \to \bar{\rho}$: average particle density.

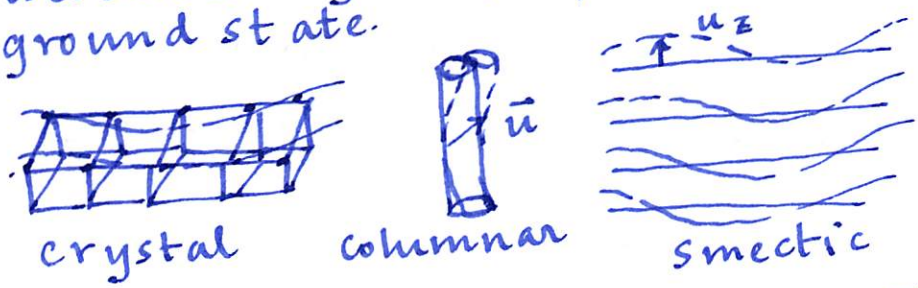
Liquid xtal: Liquid like order exists in ~~one~~ at least one space direction and atleast in one direction system is anisotropic

Types of Lxtals:

1. Imposing no positional order or imposing positional order in one or two but not three dimensions.
   a) No positional order, orientational order along a preferred axis - nematic.
   b) one dimensional order in three dimensions - smectics
   c) Two-dimensional order in 3-dimensions - liquid tubes - columnar phases.
   d) Nematic + chiral molecule → helical distortion similar to cholesterol esters → cholesteric.

Long range quasilong range order in Lxtals:

- Small deformations of crystals and liquids crystals described by $\vec{u}$ (displacement field) of lattice from its ground state.



$$F_{crystal} = F_{cr} = \frac{c}{2} \int d^3x \, (\vec{\nabla} u)^2$$

$$F_{columnar} = F_{co} = \frac{c}{2} \int d^3x \left[(\vec{\nabla}_\perp \vec{u})^2 + \lambda^2 (\vec{\nabla}_z^2 \vec{u})^2 \right]$$

$$F_{smectic} = F_s = \frac{c}{2}\int d^3x \left[(\vec{\nabla}_z \vec{u})^2 + \lambda^2 (\vec{\nabla}_\perp^2 u)^2\right]$$

Fourier transforming one can write

$$F_{cr} = \frac{c}{2}\int \frac{d^3q}{(2\pi)^3} q^2 |u(q)|^2$$

$$F_{co} = \frac{c}{2}\int \frac{d^3q}{(2\pi)^3} (q_\perp^2 + \lambda^2 q_z^4) |u(q)|^2$$

$$F_s = \frac{c}{2}\int \frac{d^3q}{(2\pi)^3} (q_z^2 + \lambda^2 q_\perp^4) |u(q)|^2.$$

From equipartition theorem. $\langle |u(q)|^2 \rangle$ each mode carries $\frac{1}{2} k_B T$.

$$\langle |u(q)|^2 \rangle = \frac{k_B T}{c q^2} \text{ - crystals.}$$

$$\langle |u(q)|^2 \rangle = k_B T / c(q_\perp^2 + \lambda^2 q_z^4) \text{ - columnar}$$

$$\langle |u(q)|^2 \rangle = k_B T / c(q_z^2 + \lambda^2 q_\perp^4) \text{ - smectics.}$$

$$\langle u^2(\vec{r}) \rangle = \frac{k_B T}{(2\pi)^3 c}\int_{L^{-1}}^{q_c} \frac{d^3q}{q^2} \cong \frac{k_B T q_c}{2\pi^2 c} \quad \text{xtals.}$$

$$\langle u^2(\vec{r}) \rangle = \frac{k_B T}{(2\pi)^3 c}\int_{L^{-1}}^{q_c} \frac{d^3q}{q_\perp^2 + \lambda^2 q_z^4} \cong k_B T (q_c \lambda^{-1})^{1/2} / 2\pi^2 c$$

- columnar.

$$\langle u^2(\vec{r}) \rangle = \frac{k_B T}{(2\pi)^3 c}\int_{L^{-1}}^{q_c} \frac{d^3q}{(q_z^2 + \lambda^2 q_\perp^4)} \cong k_B T \ln(q_c L) / 4\pi\lambda c.$$

- smectics.

$\langle u^2(\vec{r}) \rangle_{cr,\,co} \to$ finite as $L \to \infty$ controlled by short range forces.

$\langle u^2(\vec{r}) \rangle_{smectic} \to \infty$ as $L \to \infty$ fluctuations destroy long-range order: Landau-Peierls instability.

Order & Order parameter:

Nematic order parameter: $Q_{ij} = \frac{V}{N}\sum_\alpha \left(v_i^\alpha v_j^\alpha - \frac{1}{3}\delta_{ij}\right)$ $\delta(\vec{x}-\vec{x}^\alpha)$

Q_{ij} : Rank two traceless symmetric tensor.

$$\langle \underline{Q} \rangle = \begin{pmatrix} \frac{2}{3}S & 0 & 0 \\ 0 & -\frac{1}{3}S+\eta & 0 \\ 0 & 0 & -\frac{1}{3}S-\eta \end{pmatrix} ; \quad \eta \neq 0 \ \underline{Q} = \text{biaxial.} \quad \eta = 0 \ \underline{Q} = \text{uniaxial}$$

$\langle Q_{ij} \rangle = S\left(n_i n_j - \frac{1}{3}\delta_{ij}\right)$; $\hat{n}$ = Frank director.

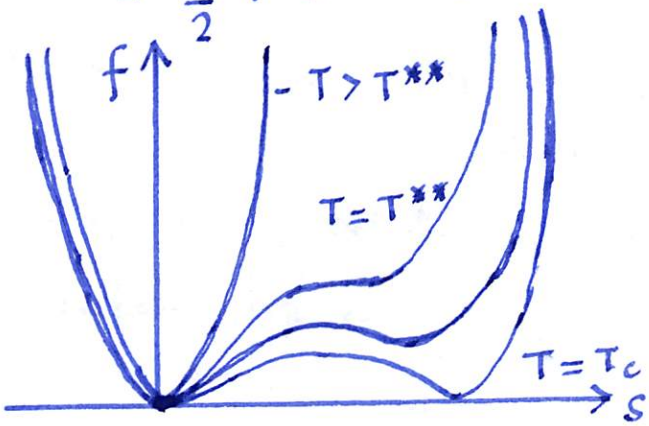
$S = \frac{1}{2}\left(3(v^\alpha . \hat{n})^2 - 1\right)^* = \frac{1}{2}\langle (3\cos^2\theta^\alpha - 1)\rangle$

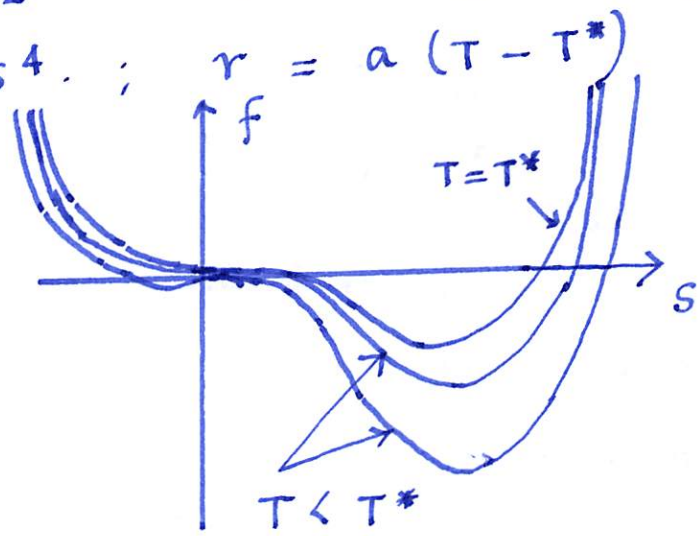
Landau Theory of nematic isotropic transition.

Phenomenological free energy functional in powers of $(Tr\langle Q\rangle^p)$

$$f = \frac{1}{2} r\left(\frac{3}{2} Tr\langle \underline{Q}\rangle^2\right) - w\left(\frac{9}{2} Tr\langle \underline{Q}\rangle^3\right) + u\left(\frac{3}{2} Tr\langle \underline{Q}\rangle^2\right)^2$$

$$= \frac{1}{2} r s^2 - w s^3 + u s^4 . \quad ; \quad r = a(T - T^*)$$





$$\frac{\partial f}{\partial s} = 0 = (r - 3ws + 4us^2)s = 0$$

$$f = \left(\frac{1}{2} r - ws + us^2\right) s^2 = 0$$

$$s_c = \frac{w}{2u} \quad ; \quad r_c = a(T_c - T^*) = \frac{w^2}{2u}$$

Latent heat $f = \frac{1}{2}(r - r_c) s_c^2 = \frac{1}{2}(r - r_c)\left(\frac{w}{2u}\right)^2$

Entropy density $= s = -\frac{\partial f}{\partial T} = -\frac{1}{2} a s_c^2 = -\frac{1}{2} a \left(\frac{w}{2u}\right)^2$

q = latent heat $= -T_c s = \frac{1}{2} a T_c \left(w/2u\right)^2$

Anisotropic molecules are anisotropic diamagnets.

$$H_{ext} = -\int d^d x \ \chi_a Q_{ij} H_i H_j = -\frac{3}{2}\int d^d x \ \chi_a H^2 S.$$

$h = (3/2)\chi_a H^2$ field conjugate to S.

Susceptibility associated with s.

$$\chi = \frac{\partial s}{\partial h} = (r - 6ws + 12us^2)^{-1}$$

Limit of metastability of nematic phase.

$\chi^{-1} = 0 \quad r^{**} = a(T^{**} - T^{*}) = 9w^2/16u.$

Elastic distortions of liquid crystals: continuum theory

Variations of $\vec{n}(\vec{r})$ small on molecular scale.

$a \nabla n \ll 1.$

$F_d =$ (i) Elastic distortion energy must be even in $\vec{n}$

(ii) no terms linear in $\vec{\nabla}\vec{n}$. Only terms invariant under rotations $(\vec{\nabla}.n)$ (ruled out by (i))

$\vec{n}.(\vec{\nabla}\times\vec{n})$ changes sign $x \to -x$, $y \to -y$, $z \to -z$. not present for nematics.

(iii) Terms of the form $\mathrm{div}(\vec{u}(\vec{r}))$; $\vec{u}(\vec{r})$ arbitrary vector field are disregarded.

Construct F_d from derivatives of $\vec{n}(\vec{r})$. Define strain $e_{\alpha\beta} = \frac{1}{2}[\partial_\alpha n_\beta + \partial_\beta n_\alpha]$ → symmetric part.

Antisymmetric part $(\vec{\nabla}\times\vec{n})_z = \frac{\partial}{\partial x}\vec{n}_y - \frac{\partial}{\partial y}\vec{n}_x$

Go to frame where z axis parallel to $\vec{n}$. All gradients of n_z vanish.

$$\nabla(n_x^2 + n_y^2 + n_z^2) = 0. \Rightarrow 2n_z \nabla n_z = 0.$$

$$e_{zz} = 0 \quad e_{zx} = \frac{1}{2}(\vec{\nabla}\times\vec{n})_y \quad e_{zy} = -\frac{1}{2}(\vec{\nabla}\times\vec{n})_x$$

$$e_{xx} + e_{yy} + e_{zz} = \vec{\nabla}\cdot\vec{n} = (e_{xx} + e_{yy})$$

$$F_d = F_e + F_c + F_{ec}$$

$$F_e = \lambda_1 e_{zz}^2 + \lambda_2 (e_{xx} + e_{yy})^2 + \lambda_3 e_{\alpha\beta} e_{\beta\alpha} + \lambda_4 e_{zz}(e_{xx} + e_{yy}) + \lambda_5 (e_{xz}^2 + e_{yz}^2).$$

F_c terms quadratic in $(\vec{\nabla} \times \vec{n})$

$$F_c = \mu_1 (\vec{\nabla} \times \vec{n})_z^2 + \mu^2 \{ (\nabla \times \vec{n})_x^2 + (\vec{\nabla} \times \vec{n})_y^2 \}$$

$$= \mu_1 [\vec{n} \cdot (\vec{\nabla} \times n)]^2 + \mu_2 [\vec{n} \times (\vec{\nabla} \times \vec{n})]^2$$

$$F_{ec} = \text{cross term} = \nu [\vec{n} \times (\vec{\nabla} \times \vec{n})]^2$$

$$\boxed{F = \frac{1}{2} K_1 (\vec{\nabla} \cdot \vec{n})^2 + \frac{1}{2} K_2 [\vec{n} \cdot (\vec{\nabla} \times \vec{n})]^2 + \frac{1}{2} K_3 [\vec{n} \times (\vec{\nabla} \times \vec{n})]^2}$$

Frank Free energy:

Simplifies if $K_1 = K_2 = K_3 = K$. $F_d = \frac{1}{2} K [(\vec{\nabla} \cdot \vec{n})^2 + (\vec{\nabla} \times \vec{n})^2]$

Tensorially written as $F_d = \frac{1}{2} K \partial_\alpha n_\beta \partial_\alpha n_\beta$.

Equilibrium conditions: $\mathcal{F}_d = \int F_d \, d\vec{r}$ minimum for all variations of $\vec{n}(\vec{r})$ which keep $\vec{n}^2 = 1$. Introduce Lagrange multiplier to take this constraint into account.

$$\delta \mathcal{F}_d = \int d\vec{r} \, \frac{1}{2} \lambda(\vec{r}) \, \delta(n^2) = \int d\vec{r} \, \lambda(\vec{r}) \, \vec{n} \cdot \delta n(\vec{r})$$

Noting F_d quadratic function of $g_{\alpha\beta} = \partial_\alpha n_\beta$.

$$\delta \mathcal{F}_d = \int \left\{ \frac{\partial F_d}{\partial n_\beta} \delta n_\beta + \frac{\partial F_d}{\partial g_{\alpha\beta}} \partial_\alpha (\delta n_\beta) \right\} d\vec{r}$$

Integrate second term by parts

$$\delta \mathcal{F}_d = \int \left\{ \frac{\partial F_d}{\partial n_\beta} - \partial_\alpha \left(\frac{\partial F_d}{\partial g_{\alpha\beta}} \right) \right\} \delta n_\beta \, d\vec{r}$$

$$h_\beta \equiv - \frac{\partial F}{\partial n_\beta} + \partial_\alpha \left(\frac{\partial F}{\partial g_{\alpha\beta}} \right) = - \lambda(\vec{r}) \, n_\beta$$

↳ Molecular field.

$$\vec{h} = h_s + h_T + h_B$$

$$\vec{h}_s = K_1 \vec{\nabla} (\vec{\nabla} \cdot \vec{n})$$

$$\vec{h}_T = -K_2 \{ \vec{A} (\vec{\nabla} \times \vec{n}) + \vec{\nabla} \times (\vec{A} \vec{n}) \}$$

$$\vec{h}_B = K_3 \{ \vec{B} \times (\vec{\nabla} \times \vec{n}) + \vec{\nabla} \times (\vec{n} \times \vec{B}) \}$$

$$\vec{A} = \vec{n} \cdot (\vec{\nabla} \times \vec{n}). \quad \vec{B} = \vec{n} \times (\vec{\nabla} \times \vec{n})$$

In one constant approx. $\boxed{\vec{h} = K \vec{\nabla}^2 \vec{n}}$

Physics By The Lake: Problems on Soft Matter Physics

Buddhapriya Chakrabarti

August 16, 2016

1 Outline

These problems are meant elucidate the concepts taught in the lectures. The overall aim is an exposition of how to develop continuum models for soft systems that are amenable to field theoretic techniques from lattice models of the same. The first problem illustrates the connection between spin systems and polymers, mostly via the use of classical statistical mechanics. The second problem involves functional integration and connects with the section on absence of long-range order in liquid crystals. It is hoped that after solving this problem you will be able to compute director fluctuations for various ordered liquid crystalline phases outlined in the lectures. The third problem involves calculating fluctuations of semi-flexible membranes in Monge gauge. It is hoped that the problem will be instructive towards understanding height-height correlation functions for fluctuating sheets (a quantity that can be probed via scattering experiments)

2 Problems

1. *Model of DNA stretching with a force:*

   (a) Consider first the freely jointed chain model of DNA stretching in *one* dimension as shown in Fig. 1a, *i.e.*, a set of N links (labelled by integers $\{j\}$) of length a which point along $+x$ if $\sigma_j = +1$ and along $-x$ if $\sigma_j = -1$.

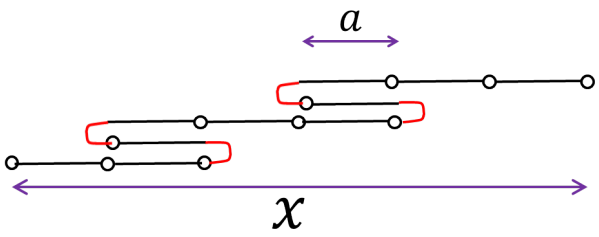


Figure 1:

Assume that the energy associated with a force f applied to the two ends of the chain is

$$E = A - fx, \tag{1}$$

where

$$x = a \sum_j \sigma_j, \tag{2}$$

and A is a constant. Use statistical mechanics to calculate the averaged stretched length $\langle x \rangle$ as a function of the applied force f and temperature T. Discuss your result in the limits $fa/k_BT << 1$ and $fa/k_BT >> 1$. Describe how these asymptotic behaviours differ, if at all, from 3-dimensional freely jointed chain (see notes).

(b) Now consider a "worm-like chain" modification of the total energy for N links, namely

$$E = B - J \sum_{\langle ij \rangle} \sigma_i \sigma_j - fx. \tag{3}$$

Here $J > 0$ represents a tendancy of neighbouring links to point in parallel direction instead of "backtracking" and pointing antiparallel and B is a constant. In the limit of zero force, calculate the "radius of gyration" $\sqrt{\langle x^2 \rangle_N}$ of this one-dimensional model. (Hint: Modify the solution for the one-dimensional Ising model in a magnetic field for *open* boundary conditions.)

(c) Use results derived in (b) to compute $\langle x \rangle$ vs. f. Sketch your result and contrast this graph with that arising from part (a). Do the high force limits (approaches to saturation) differ?

2. *Functional Integration:* Consider the statistical mechanics of a one-dimensional harmonic chain of length L, defined with respect to sites $x_m = lm$ as shown in Fig. 2.

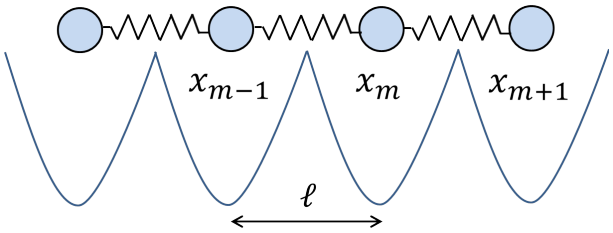


Figure 2:

Each particle is connected to its neighbour by a spring and sits in a potential well centered on x_m. Thus if an actual particle position is $r_m = x_m + u_m$, the energy is

$$E(u_1, u_2, \ldots u_N) = \frac{B'}{2} \sum_m (u_{m+1} - u_m)^2 + \frac{C'}{2} \sum_m u_m^2, \tag{4}$$

(a) Take the continuum limit of Eq. 4, to obtain an energy which depends on the function $u(x)$

$$E(u(x)) = \frac{1}{2} B \int_0^L dx \left(\frac{du(x)}{dx} \right)^2 \frac{C}{2} \int_0^L dx u^2(x). \tag{5}$$

(b) What are the constants B and C?

(c) Assume periodic boundary conditions and rewrite the energy E in terms of Fourier variables defined by

$$u(x) = \frac{1}{L} \sum_q e^{iqx} \hat{u}(q). \tag{6}$$

(d) Taking into account the constraint $\hat{u}^*(q) = \hat{u}(-q)$ (*i.e.* $u(x)$ is a real function, evaluate the Free energy F defined by a functional integral

$$e^{-F/k_B T} = \int D\hat{u}(x) e^{-E/k_B T}, \tag{7}$$

where

$$\int Du(x) = \prod_q \int_{-\infty}^{\infty} dRe\hat{u}(q) \int_{-\infty}^{\infty} dIm\hat{u}(q) \tag{8}$$

(e) What q values are taken for the product.

(f) Express F in the limit $L \to \infty$ as an integral over wavevectors q.

(g) Does the integral converge for large q?

(h) If not, what physically reasonable maximum value of q should you take.

(i) Using similar methods show that (with L finite)

$$\langle \hat{u}(q)\hat{u}(k)\rangle = \frac{k_B T}{A(q)}\delta_{q,-k} \tag{9}$$

and determine $A(q)$.

(j) Use your result to evaluate $\langle u^2(x)\rangle$ in the limit $L \to \infty$.

(k) Discuss the implications of your results in the part above for crystalline order in one-dimension in the limit $C \to 0$.

(l) Can an ordered phase exist in one dimension in this limit.

3. *Membrane fluctuations* Thermal fluctuations play an important role in lipid bi-layer membranes which are a main constituent of cell walls. Similar to worm-like polymer chains, there is a bending rigidity $\tilde{\kappa}$ which tends to keep these membranes flat. Imagine dividing the membrane into a network of square plaquettes as shown in Fig. 3.

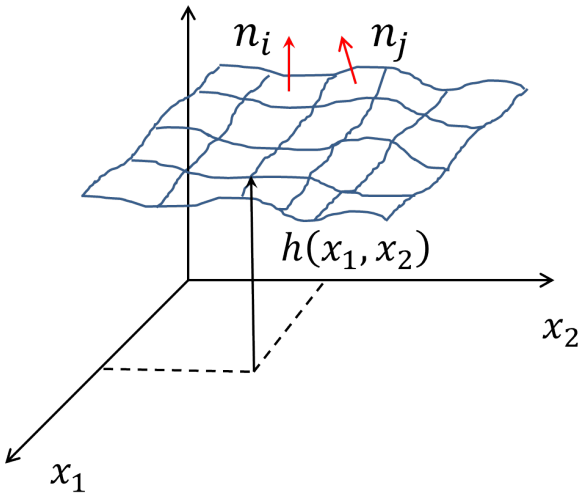


Figure 3:

Eah plaquette is characterised by a unit normal $\hat{n}_i$. The surface is paramatrised by a height function $h(x_1, x_2)$, so that a point on the surface is given by $\vec{r}(x_1, x_2) = (x_1, x_2, h(x_1, x_2))$.

(a) Show that the unit normal at point (x_1, x_2) is

$$\vec{n}(x_1, x_2) = \left(-\frac{\partial h}{\partial x_1}, -\frac{\partial h}{\partial x_2}, 1 \right) / \sqrt{1 + \left(\vec{\nabla} h\right)^2}, \tag{10}$$

(b) Take the continuum limit of an energy function

$$E = -\tilde{\kappa} \sum_{\langle ij \rangle} \vec{n}_i \cdot \vec{n}_j,$$

where the sum is over nearest neighbour pairs of plaquettes on a square lattice with lattice constant a. Use results of part (a) to obtain an expression of $E[h(x_1, x_2)]$ to leading order in an expansion in gradients f, namely

$$E \approx \frac{1}{2} \int d^d x \left[\kappa_a (\nabla^2 f)^2 - \kappa_b \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right) \right]. \tag{11}$$

What are the coupling constants κ_a and κ_b?

(c) Neglect the second term of Eq. 11 (This is to be expected since it is a total derivative and only contributes to boundaries). Explore the possibility of long-range order in the normals (*i.e.* "flat" membrane phase): Assume a non-zero $\langle \vec{n} \rangle \parallel \hat{z}$ and calculate

$$\langle \vec{n} \cdot \hat{z} \rangle = \langle \frac{1}{\sqrt{1 + |\vec{\nabla} h|^2}} \rangle \approx 1 - \frac{1}{2} \langle |\vec{\nabla} h|^2 \rangle, \tag{12}$$

using a Fourier method for worm-like chains. Assume periodic boundary conditions in x_1 and x_2 and introduce short and large distance cut offs as necessary. Define a membrane persistence length l such that the second term of Eq. 12 becomes comparable to the first. Write your result in the form $l = af(\kappa_a / k_B T)$, where a is a microscopic length. What is the function $f(x)$?

SCQ

Strongly correlated quantum systems (SCQ): lecture notes for Physics by the Lake 2018

Chris Hooley (SUPA, University of St Andrews)

first written 2nd September 2006;
last updated 25th July 2018

Contents

1 Second quantisation and the Fermi sea **2**
1.1 In two sentences... . . . 2
1.2 The harmonic oscillator as a single bosonic mode . . . 2
1.3 Many bosons . . . 4
1.4 Many fermions; the Pauli exclusion principle . . . 5
1.5 The Fermi sea; Friedel oscillations . . . 7
1.6 Exercises . . . 9

2 Electron-electron interactions: Landau Fermi liquid theory **12**
2.1 In two sentences... . . . 12
2.2 Coulomb beats kinetic: a back-of-the-envelope calculation . . . 12
2.3 A perturbative approach; phase space restrictions . . . 13
2.4 Adiabatic continuity; energy functional and Landau parameters . . . 16
2.5 Instabilities of the Fermi liquid: Stoner ferromagnetism . . . 18
2.6 Relations between the Landau parameters . . . 19
2.7 Exercises . . . 20

3 Magnetic impurities: the Anderson model and the Kondo effect **21**
3.1 In two sentences... . . . 21
3.2 Magnetic susceptibilities: a quick recap . . . 21
3.3 Why do magnetic impurities exist?: the Anderson model . . . 24
3.4 The Kondo limit . . . 25
3.5 Variational approach: the Varma-Yafet solution . . . 26
3.6 Recovery of the Fermi liquid at low temperatures . . . 29
3.7 Exercises . . . 29

4 The Hubbard model and the Mott insulator **30**
4.1 In two sentences... . . . 30
4.2 The tight-binding model . . . 30
4.3 The Hubbard model and the Mott insulator . . . 32
4.4 Magnetism in the Mott state . . . 34
4.5 Exercises . . . 35

5 Magnetism in Mott insulators: the Heisenberg model analysed **36**
5.1 In two sentences... . . . 36
5.2 Classical magnetism . . . 36
5.2.1 Fourier (or spin-wave) representation . . . 38
5.2.2 Obtaining the classical ground state(s) . . . 39
5.3 Semi-classical magnetism (spin wave theory) . . . 41
5.3.1 The Holstein-Primakoff transformation . . . 42
5.3.2 Diagonalising the Hamiltonian for a simple ferromagnet . . . 43
5.3.3 Checking the consistency of the spin wave approach . . . 44
5.4 Exercises . . . 46

6 A survey of experiment **47**
6.1 In two sentences... . . . 47
6.2 The high-temperature superconductors: doping a Mott insulator? . . . 47
6.2.1 Crystal structures . . . 47
6.2.2 Basic experimental information . . . 47
6.2.3 What should the model be? . . . 49
6.2.4 Discussion . . . 50
6.3 Heavy fermion materials: the Kondo lattice model? . . . 51
6.3.1 Crystal structures . . . 51
6.3.2 Basic experimental information . . . 51
6.3.3 The Doniach phase diagram and the Kondo lattice model . . . 52
6.3.4 Discussion . . . 55

1 Second quantisation and the Fermi sea

1.1 In two sentences...

Obviously traditional quantum mechanics with $\psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N; t)$ is not suited to describe systems where the number of particles changes with time. What we need instead is 'second quantisation', which is the basic tool for the theoretical description of quantum many-body systems, and which is introduced — with some applications — in this lecture.

1.2 The harmonic oscillator as a single bosonic mode

As the simplest example, let's consider the one-dimensional harmonic oscillator. The Hamiltonian is

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega^2\hat{x}^2, \tag{1}$$

where ω is the classical angular frequency of the oscillator[1]. The operators $\hat{p}$ and $\hat{x}$ obey the commutation relation $[\hat{x}, \hat{p}] \equiv \hat{x}\hat{p} - \hat{p}\hat{x} = i\hbar$.

We can't just read off the eigenvalues from this Hamiltonian, as it is made of two pieces that don't commute with each other. So we'll proceed in three steps: first make the operators dimensionless, then take linear combinations of them to turn $\hat{H}$ into a single term (plus a constant), then find the eigenvalues of this single term.

[1] To see this, start with Hooke's law, $F = -\kappa x$, where κ is the spring constant. On the other hand, $F = -\frac{\partial V}{\partial x}$, so that $V(x) = \frac{1}{2}\kappa x^2$. The classical solution gives oscillations with an angular frequency $\omega = \sqrt{\kappa/m}$, so that κ can be written $m\omega^2$.

- *Dimensionless operators.* The constants available are m (kg), ω (s^{-1}), and $\hbar$ (kg m^2 s^{-1}); from these we can make only one length and one momentum scale:

$$x_0 = \sqrt{\frac{\hbar}{m\omega}}; \qquad p_0 = \frac{\hbar}{x_0} = \sqrt{\hbar m\omega}. \tag{2}$$

So we define the dimensionless operators $\hat{X}$ and $\hat{P}$ by

$$\hat{x} = x_0\hat{X}; \qquad \hat{p} = p_0\hat{P}, \tag{3}$$

whence we derive their commutation relation

$$\left[\hat{X}, \hat{P}\right] = \frac{1}{x_0 p_0}\left[\hat{x}, \hat{p}\right] = i. \tag{4}$$

In terms of these new operators, the Hamiltonian is written[2]

$$\hat{H} = \frac{\hbar\omega}{2}\left(\hat{X}^2 + \hat{P}^2\right). \tag{5}$$

- *Resolving the 'two non-mutually-commuting terms' problem.* The next step is the real magic. The basic idea of it is that the combination $X^2 + P^2$ can be roughly written as $(X - iP)(X + iP)$ (up to commutation issues), which leads us to define the following operator:

$$\hat{a} \equiv \frac{1}{\sqrt{2}}\left(\hat{X} + i\hat{P}\right). \tag{6}$$

This operator (called an **annihilation operator** — we'll see why below) is not Hermitian; its conjugate is

$$\hat{a}^\dagger = \frac{1}{\sqrt{2}}\left(\hat{X} - i\hat{P}\right), \tag{7}$$

which is a **creation operator**. From (4) it's easy to show that

$$\left[\hat{a}, \hat{a}^\dagger\right] = 1. \tag{8}$$

In terms of these operators, the Hamiltonian becomes

$$\hat{H} = \hbar\omega\left(\hat{a}^\dagger\hat{a} + \frac{1}{2}\right). \tag{9}$$

- *Finding the eigenvalues.* The point now is that we can *completely forget* where the creation and annihilation operators came from, and just treat them as objects obeying their own algebra (8). Here's a pseudo-derivation of the eigenvalues on that basis[3].

We conjecture a state $|0\rangle$ that is destroyed by the annihilation operator[4]:

$$a\,|0\rangle = 0. \tag{10}$$

[2] Notice the pleasing symmetry between the momentum and position operators.

[3] For a much more respectable treatment, see for example C. Cohen-Tannoudji, B. Diu, and F. Laloë, *Quantum Mechanics* (Wiley, 1977).

[4] I'm going to stop putting hats on these operators now. When you see an a or an $a^\dagger$ in a formula henceforth, it's an operator.

We call this state the **vacuum**. Hence, multiplying from the left by $a^\dagger$, we get

$$a^\dagger a \left|0\right\rangle = 0, \tag{11}$$

i.e. the vacuum is an eigenstate of $a^\dagger a$ with eigenvalue 0. We call $a^\dagger a$ the **number operator**. To make a new state from $|0\rangle$ we can't use a (because of (10)), so we must use $a^\dagger$:

$$\left|1\right\rangle \equiv a^\dagger \left|0\right\rangle . \tag{12}$$

If we act the number operator on this new state, we get

$$a^\dagger a \left|1\right\rangle = a^\dagger a a^\dagger \left|0\right\rangle = a^\dagger \left(1 + a^\dagger a\right) \left|0\right\rangle = a^\dagger \left|0\right\rangle + a^\dagger \underbrace{\left(a^\dagger a \left|0\right\rangle\right)}_{=0 \text{ by } (11)} = \left|1\right\rangle , \tag{13}$$

where we have used the commutation relation (8) in the second step; so $|1\rangle$ is also an eigenstate of the number operator, with eigenvalue 1. It can be shown similarly that the spectrum of possible eigenvalues is just the integers from 0 upwards (see Exercises), and hence the eigenvalues of $\hat{H}$ are given by

$$E_n = \hbar\omega \left(n + \frac{1}{2}\right), \qquad n = 0, 1, 2, \ldots \tag{14}$$

Let's pause for a moment to interpret that result. In the 'first-quantised' picture of the simple harmonic oscillator, there is always one particle (of mass m, attached to the spring). It has certain energies that arise from the quantisation of its motion. In the 'second-quantised' picture, we think of these quanta of energy as particles in their own right. Thus the state $|0\rangle$ corresponds to an absence of particles, the state $|1\rangle$ to the presence of one particle, and so forth. Since we can have as many particles as we like[5], we interpret them as bosons, and thus take (8) to be the **fundamental commutation relation for bosonic operators**.

Very many problems in condensed matter theory are solved by forcing the Hamiltonian into the (usually approximate) form of a sum of number operators. This operation is termed **diagonalising** the Hamiltonian. Then, because we know the possible eigenvalues of the number operators, we know the possible eigenvalues of the Hamiltonian. Armed with that energy spectrum, we can calculate the system's thermodynamics[6].

1.3 Many bosons

The generalisation of the above considerations to the case of many species of boson is fairly straightforward[7]. These different species of boson might represent different modes of a bosonic field (a_j annihilating a quantum of energy in mode j), or bosons on different sites of a crystal lattice (a_j annihilating a boson on site j). We shall be mainly interested in the latter example.

The basic property of (indistinguishable) bosons is that their wavefunction is symmetric under interchange of particle labels, so we want to guarantee that

$$\left|i, j\right\rangle = \left|j, i\right\rangle . \tag{15}$$

[5] Equivalently, there is no state $|x\rangle$ for which $a^\dagger |x\rangle = 0$.

[6] At least at low temperatures. Most of the approximations that we make to force the Hamiltonian into number-operator form get worse as the temperature increases.

[7] Though again my presentation here is rather slapdash. To see it done properly, consult for example J.W. Negele and H. Orland, *Quantum many-particle systems* (Addison-Wesley, 1988).

Since these states will be made by creating the bosons from the vacuum, this is the same as the statement that

$$a_i^\dagger a_j^\dagger |0\rangle = a_j^\dagger a_i^\dagger |0\rangle , \tag{16}$$

i.e. the creation operators on different sites *commute*. This is clearly also true of the annihilation operators, and furthermore of any combination of creation and annihilation operators that refer to different sites. Combining this observation with (8), we obtain[8]

$$\begin{aligned} \left[a_j^\dagger, a_k^\dagger\right] &= 0; \\ [a_j, a_k] &= 0; \\ \left[a_j, a_k^\dagger\right] &= \delta_{jk}. \end{aligned} \tag{17}$$

These are the **fundamental commutation relations for many species of boson**. A general state is described by specifying the number of bosons in each mode, and is created from the vacuum by acting each creation operator the appropriate number of times (for normalisation see Exercises):

$$|n_1, n_2, \ldots, n_k, \ldots\rangle = \frac{\left(a_1^\dagger\right)^{n_1}}{\sqrt{n_1!}} \frac{\left(a_2^\dagger\right)^{n_2}}{\sqrt{n_2!}} \ldots \frac{\left(a_k^\dagger\right)^{n_k}}{\sqrt{n_k!}} \ldots |0\rangle . \tag{18}$$

It is important to remember, when using bosonic operators, that they don't usually preserve the normalisation of the state — see the Exercises for an example of this.

A simple example of a Hamiltonian involving many bosonic species is the Hamiltonian for photons in a cavity:

$$H = \sum_{\mathbf{k}\eta} \varepsilon_{\mathbf{k}} a_{\mathbf{k}\eta}^\dagger a_{\mathbf{k}\eta}. \tag{19}$$

Here each mode is identified by its wavevector $\mathbf{k}$, where $\mathbf{k}$ takes all values consistent with the boundary conditions imposed by the cavity (which may be periodic or hard-wall), and its polarisation η, which takes one of two values (e.g. horizontal and vertical)[9]. The energy of mode $\mathbf{k}\eta$ is

$$\varepsilon_{\mathbf{k}} = \hbar c |\mathbf{k}|, \tag{20}$$

independently of its polarisation η. The relation (20) is called the **dispersion relation** for free photons. Since the Hamiltonian (19) is a sum of number operators, we know that its eigenvalues are

$$E\left(\{n_{\mathbf{k}\eta}\}\right) = \sum_{\mathbf{k}\eta} \varepsilon_{\mathbf{k}} n_{\mathbf{k}\eta}, \tag{21}$$

where each $n_{\mathbf{k}\eta}$ is an independently chosen (zero or positive) integer. Armed with this energy spectrum, we can obtain the thermodynamic properties of a gas of photons at temperature T (see Exercises).

1.4 Many fermions; the Pauli exclusion principle

The extension of the second-quantised scheme to fermions works similarly, except that this time we need to incorporate the fact that the many-fermion wavefunction is *antisymmetric* under

[8] Here, δ_{jk} is the Kronecker delta symbol, which takes the value 1 when $j = k$ and 0 otherwise.

[9] The fact that the photon, as a spin-1 particle, has two rather than three polarisation states may come as a surprise. The resolution comes from special relativity: one of the polarisation states of the photon is suppressed because it is a zero-mass particle, and thus cannot be brought to rest.

interchange of particle labels. In this case, therefore,

$$|i,j\rangle = -\,|j,i\rangle\,, \tag{22}$$

which gets translated into creation-operator notation as[10]

$$c_i^\dagger c_j^\dagger\,|0\rangle = -c_j^\dagger c_i^\dagger\,|0\rangle\,. \tag{23}$$

This means that fermionic creation operators **anticommute**; thus, in the basic relations analogous to (17), we replace the commutator by the anticommutator[11]:

$$\begin{aligned} \left\{c_j^\dagger, c_k^\dagger\right\} &= 0; \\ \{c_j, c_k\} &= 0; \\ \left\{c_j, c_k^\dagger\right\} &= \delta_{jk}. \end{aligned} \tag{24}$$

As before, there is a vacuum state $|0\rangle$, containing no particles, which satisfies

$$c_k\,|0\rangle = 0 \qquad \forall k. \tag{25}$$

The Pauli exclusion principle is a trivial result of the anticommutation properties of the creation operators. To see this, we note that a state with two particles in the same mode k would be written

$$|2_k\rangle = c_k^\dagger c_k^\dagger\,|0\rangle\,. \tag{26}$$

But the fundamental anticommutation relations (24) tell us that

$$\left\{c_k^\dagger, c_k^\dagger\right\} = c_k^\dagger c_k^\dagger + c_k^\dagger c_k^\dagger = 2c_k^\dagger c_k^\dagger = 0, \tag{27}$$

and hence the state $|2_k\rangle$ does not exist[12]. Thus the fermionic number operator

$$n_k = c_k^\dagger c_k \tag{28}$$

has only two eigenvalues, 0 or 1.

In the same way that the normalisation factors have to be looked after carefully when dealing with many-boson systems, the minus signs in many-fermion systems must be kept under control. In a Fermi system, it is no longer sufficient to supply simply the occupation numbers of the various modes. To see this, consider a two-mode system with both modes occupied. This could be created in one of two ways:

$$|1,1\rangle_1 = c_1^\dagger c_2^\dagger\,|0\rangle \qquad \text{or} \qquad |1,1\rangle_2 = c_2^\dagger c_1^\dagger\,|0\rangle\,; \tag{29}$$

since the creation operators anticommute, $|1,1\rangle_1 = -\,|1,1\rangle_2$. These minus signs may be very important, since they can cause quantum interference which is at the root of many physical phenomena; an example follows in the next section. What this means in practice is that a state is defined by its occupation numbers only if there is also a convention for the *order* in which they are to be populated. A useful rule in practical calculations is to try and manipulate the sequence of operators so that it becomes a product of number operators — their action on a given state is always unambiguous.

[10] It has become standard to reserve the letters a and b for bosonic operators, using c, d, or f for fermionic ones. I shall follow this convention.

[11] The anticommutator of two operators is defined by $\{A, B\} \equiv AB + BA$.

[12] Another way to view this difference between fermions and bosons is that every operator commutes with itself — hence no exclusion principle in the bosonic case — whereas for an operator to anticommute with itself is a special requirement.

1.5 The Fermi sea; Friedel oscillations

A very common condensed matter application of the many-body theory of fermions is to describe the sea of conduction electrons in a metal. In the simplest model, due to Sommerfeld, the 'metal' is viewed as an empty cubic box of side length L which contains the electrons. Thus, the available modes for the electrons are dictated by the boundary conditions of the box, leading to a discrete set of allowed $\mathbf{k}$ values. We can think of these $\mathbf{k}$-modes as a normal-mode expansion of the electron field operator[13]:

$$\psi_\sigma(\mathbf{r}) = \frac{1}{\sqrt{V}} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} c_{\mathbf{k}\sigma}. \tag{30}$$

Here, σ is a spin label (taking the value $\uparrow$ or $\downarrow$), $V = L^3$ is the volume of the box, and we have assumed periodic boundary conditions at the box's edges. The mode label thus now consists of a momentum label $\mathbf{k}$ and a spin label σ. The total density operator (counting both spin species) is given by

$$\begin{aligned} \rho(\mathbf{r}) &= \sum_\sigma \psi_\sigma^\dagger(\mathbf{r})\psi_\sigma(\mathbf{r}) \\ &= \frac{1}{V} \sum_{\mathbf{k}\mathbf{q}\sigma} e^{i\mathbf{r}\cdot(\mathbf{q}-\mathbf{k})} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{q}\sigma}. \end{aligned} \tag{31}$$

The Hamiltonian of the Fermi gas is[14]

$$H = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma}, \tag{32}$$

where the dispersion relation for free particles of mass m is given by[15]

$$\varepsilon_{\mathbf{k}} = \frac{\hbar^2 \mathbf{k}^2}{2m}. \tag{33}$$

Because of the exclusion principle, the ground state for N fermions consists of placing them in pairs (spin up and spin down) in each of the $N/2$ lowest-energy states. Because the dispersion relation (33) depends only on $|\mathbf{k}|$, these define a sphere in $\mathbf{k}$-space: the **Fermi sphere**. We call this ground state $|FS\rangle$.

As an example of a second-quantised calculation, let us obtain the density-density correlation function in the ground state[16]:

$$C(\mathbf{r}) \equiv \langle FS|\, \rho(\mathbf{r})\rho(\mathbf{0})\, |FS\rangle - \left(\langle FS|\, \rho(\mathbf{0})\, |FS\rangle\right)^2. \tag{34}$$

This is related to the conditional probability that there is an electron at position $\mathbf{r}$ given that there is one at position $\mathbf{0}$; it is of experimental relevance mainly for STM[17] and other tunnelling

[13] The electron field operator $\psi_\sigma(\mathbf{r})$ annihilates an electron with spin σ at position $\mathbf{r}$.

[14] Since there are no applied magnetic fields, the energy is independent of the spin projection σ.

[15] The true dispersion relation in solids is always different from this, because of the effects of the periodic potential due to the crystal lattice. Here, we completely neglect the lattice potential; you might want to think about why the resulting theory often works so well.

[16] Since the system is translationally invariant in real space, I have put some of the position indices to $\mathbf{0}$ without loss of generality.

[17] Scanning Tunnelling Microscope.

probes. To calculate $C(\mathbf{r})$, we substitute in two copies of the expression for the density operator (31):

$$C(\mathbf{r}) = \frac{1}{V^2} \sum_{\mathbf{kqps}\sigma\tau} e^{i\mathbf{r}\cdot(\mathbf{q}-\mathbf{k})} \underbrace{\langle FS| c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{q}\sigma} c^\dagger_{\mathbf{p}\tau} c_{\mathbf{s}\tau} |FS\rangle}_{\equiv M} - \frac{N^2}{V^2}, \tag{35}$$

where we have used the fact that the average density of the gas is N/V.

The matrix element M is clearly zero unless the two newly populated modes match the two depopulated ones, but this can happen in two ways: either $\mathbf{k} = \mathbf{q}$ and $\mathbf{p} = \mathbf{s}$, or $\mathbf{k} = \mathbf{s}$ and $\mathbf{q} = \mathbf{p}$. In the latter case, the spins τ and σ must also be equal. Thus one might think that the matrix element would be written

$$M = \delta_{\mathbf{kq}}\delta_{\mathbf{ps}} \underbrace{\langle FS| c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} c^\dagger_{\mathbf{p}\tau} c_{\mathbf{p}\tau} |FS\rangle}_{\equiv M_H} + \delta_{\mathbf{ks}}\delta_{\mathbf{qp}}\delta_{\sigma\tau} \underbrace{\langle FS| c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{q}\sigma} c^\dagger_{\mathbf{q}\sigma} c_{\mathbf{k}\sigma} |FS\rangle}_{\equiv M_F}, \tag{36}$$

where the first term, M_H, is called the **Hartree term**, and the second, M_F, is called the **Fock (or exchange) term**.

This is not quite right, however, since both the first and second terms on the right-hand side include the term in which all four momenta are the same and the spins are parallel. To correct that double-counting, it must be subtracted again, meaning that the correct expression is

$$M = \delta_{\mathbf{kq}}\delta_{\mathbf{ps}} M_H + \delta_{\mathbf{ks}}\delta_{\mathbf{qp}}\delta_{\sigma\tau} M_F - \delta_{\sigma\tau}\delta_{\mathbf{kq}}\delta_{\mathbf{kp}}\delta_{\mathbf{ks}} \langle FS| c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} |FS\rangle . \tag{37}$$

We shall calculate each of these three terms in turn.

- The Hartree term is simple, since its operator is a product of number operators already. Let us define the symbol $\Theta_{\mathbf{k}}$ to be 1 if the state $\mathbf{k}$ is occupied in the ground state (i.e. lies within the Fermi sphere), and 0 otherwise. Using this symbol, we have that

$$c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} c^\dagger_{\mathbf{p}\tau} c_{\mathbf{p}\tau} |FS\rangle = c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} \Theta_{\mathbf{p}} |FS\rangle = \Theta_{\mathbf{k}}\Theta_{\mathbf{p}} |FS\rangle , \tag{38}$$

  whence, assuming that the state $|FS\rangle$ is normalised, we find that

$$M_H = \Theta_{\mathbf{k}}\Theta_{\mathbf{p}}. \tag{39}$$

- In the case of the Fock term, the operator to be averaged is not a product of number operators. But we may make it so using the anticommutation relations (24):

$$\begin{aligned} c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{q}\sigma} c^\dagger_{\mathbf{q}\sigma} c_{\mathbf{k}\sigma} &= c^\dagger_{\mathbf{k}\sigma} \left(1 - c^\dagger_{\mathbf{q}\sigma} c_{\mathbf{q}\sigma}\right) c_{\mathbf{k}\sigma} \\ &= c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} - c^\dagger_{\mathbf{k}\sigma} c^\dagger_{\mathbf{q}\sigma} c_{\mathbf{q}\sigma} c_{\mathbf{k}\sigma} \\ &= c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} + c^\dagger_{\mathbf{q}\sigma} c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{q}\sigma} c_{\mathbf{k}\sigma} \\ &= c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} + c^\dagger_{\mathbf{q}\sigma} \left(\delta_{\mathbf{kq}} - c_{\mathbf{q}\sigma} c^\dagger_{\mathbf{k}\sigma}\right) c_{\mathbf{k}\sigma} \\ &= c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} + \delta_{\mathbf{kq}} c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} - c^\dagger_{\mathbf{q}\sigma} c_{\mathbf{q}\sigma} c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma}. \end{aligned} \tag{40}$$

  This final expression is made entirely of number operators, and so its action on the state $|FS\rangle$ may be easily written down; thus we obtain for the Fock term

$$M_F = \Theta_{\mathbf{k}} + \delta_{\mathbf{kq}}\Theta_{\mathbf{k}} - \Theta_{\mathbf{q}}\Theta_{\mathbf{k}}. \tag{41}$$

- Finally, the correction term. The expectation value of the operators in this case is simple:

$$\langle FS|\, c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma}\, |FS\rangle = \Theta^2_{\mathbf{k}} = \Theta_{\mathbf{k}}, \tag{42}$$

since $\Theta_{\mathbf{k}} = 0, 1$. The result of this is simply to cancel out the $\delta_{\mathbf{kq}}\Theta_{\mathbf{k}}$ in the above expression for M_F.

Substituting these three terms into (37), we obtain for the full matrix element

$$M = \delta_{\mathbf{kq}}\delta_{\mathbf{ps}}\Theta_{\mathbf{k}}\Theta_{\mathbf{p}} + \delta_{\mathbf{ks}}\delta_{\mathbf{qp}}\delta_{\sigma\tau}\left(\Theta_{\mathbf{k}} - \Theta_{\mathbf{q}}\Theta_{\mathbf{k}}\right), \tag{43}$$

and hence the full density-density correlator is

$$\begin{aligned} C(\mathbf{r}) &= \frac{1}{V^2}\sum_{\mathbf{kqps}\sigma\tau} e^{i\mathbf{r}\cdot(\mathbf{q}-\mathbf{k})}\left[\delta_{\mathbf{kq}}\delta_{\mathbf{ps}}\Theta_{\mathbf{k}}\Theta_{\mathbf{p}} + \delta_{\mathbf{ks}}\delta_{\mathbf{qp}}\delta_{\sigma\tau}\left(\Theta_{\mathbf{k}} - \Theta_{\mathbf{q}}\Theta_{\mathbf{k}}\right)\right] - \frac{N^2}{V^2} \\ &= \frac{4}{V^2}\left(\sum_{\mathbf{k}}\Theta_{\mathbf{k}}\right)^2 + \frac{2}{V^2}\sum_{\mathbf{kq}} e^{i\mathbf{r}\cdot(\mathbf{q}-\mathbf{k})}\left(\Theta_{\mathbf{k}} - \Theta_{\mathbf{q}}\Theta_{\mathbf{k}}\right) - \frac{N^2}{V^2} \\ &= \left(\frac{N}{V}\right)^2 + \frac{N}{V}\delta(\mathbf{r}) - \frac{2}{V^2}\left|\sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}}\Theta_{\mathbf{k}}\right|^2 - \frac{N^2}{V^2} \\ &= \frac{N}{V}\delta(\mathbf{r}) - \frac{2}{V^2}\left|\sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}}\Theta_{\mathbf{k}}\right|^2, \end{aligned} \tag{44}$$

where some standard identities involving delta-functions have been used.

Let us pause again for interpretation. The first term of (44) is an anomalous contribution that arises only if we try to measure the correlator exactly at the origin. The second term therefore contains all the interesting physics, and we observe that it oscillates as a function of $\mathbf{r}$: for a rigorous proof of this in one dimension, see the Exercises. These oscillations are called **Friedel oscillations**. The Friedel oscillations created around impurities are sometimes a valuable way to inspect, in real space, the momentum-space distribution of the electrons in a sample. An example of the Friedel oscillations observed around impurities and defects on a beryllium surface is shown in Fig. 1.

1.6 Exercises

1. *The simple harmonic oscillator.*

   (a) Transforming the equation (10) back into the $(\hat{x}, \hat{p})$ basis, and using the usual representation

   $$\hat{x} \to x, \qquad \hat{p} \to -i\hbar\frac{\partial}{\partial x}, \tag{45}$$

   obtain the ground-state wavefunction of the simple harmonic oscillator.

   (b) Show that the state (12) is normalised.

   (c) Show (using second-quantised notation) that the normalised expression for the nth excited state is

   $$|n\rangle = \frac{\left(a^\dagger\right)^n}{\sqrt{n!}}\,|0\rangle\,. \tag{46}$$

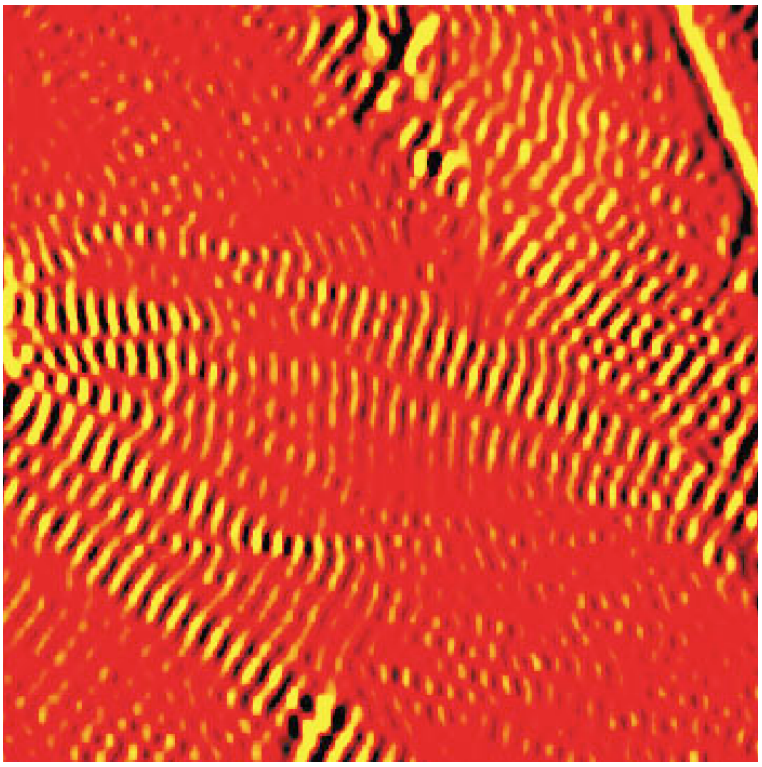


Figure 1: STM image: Friedel oscillations of electrons at the surface of a beryllium sample. Reproduced from P.T. Sprunger, L. Petersen, E.W. Plummer, E. Laegsgaard, and F. Besenbacher, *Science* **275**, 1764 (1997).

(d) Using (46) and the representation (45), obtain the spatial wavefunctions of the first and second excited states.

(e) Calculate the average $\langle 0| aaa^\dagger a^\dagger |0\rangle$. How would you interpret your result?

2. *A two-mode bosonic system.* Consider a bosonic system with two modes, and call their occupation numbers n_1 and n_2. The system is initially in state $|3,5\rangle$ (in the notation of (18)). What is the operator that moves one boson from the second mode to the first (preserving normalisation)? What is the resulting state? (*Hint:* If in doubt, write out the state $|3,5\rangle$ explicitly and use the fundamental commutation relations.)

3. *Black-body radiation.* The average number of bosons in a bosonic mode of energy ε at temperature T is given by the Bose-Einstein distribution:

$$\langle n_\varepsilon \rangle = \frac{1}{\exp\left(\frac{\varepsilon-\mu}{k_B T}\right) - 1}, \tag{47}$$

where μ is the chemical potential.

(a) Use this to calculate the average energy, at temperature T, of the photon system whose Hamiltonian is (19). (Recall that for a gas of photons in an oven, or indeed for any set of bosons whose number is not conserved, $\mu = 0$. To determine the allowed values of $\mathbf{k}$, take the cavity to be a cubic box of side-length L, and assume hard-wall boundary conditions. You may quote the standard integral

$$\int_0^\infty \frac{q^3 dq}{e^q - 1} = \frac{\pi^4}{15}. \ )$$

(b) Hence obtain the specific heat capacity of the photon gas,

$$C_V \equiv \left(\frac{\partial \langle E \rangle}{\partial T}\right)_V. \tag{48}$$

4. *Friedel oscillations.*

(a) Given the fundamental anticommutation relations (24) and the expression for the electron field operator (30), prove that

$$\left\{\psi_\sigma(\mathbf{r}), \psi_\tau^\dagger(\mathbf{r}')\right\} = \delta_{\sigma\tau}\delta^{(3)}(\mathbf{r} - \mathbf{r}'), \tag{49}$$

where the three-dimensional delta-function is defined by

$$\delta^{(3)}(\mathbf{r}) \equiv \delta(x)\delta(y)\delta(z). \tag{50}$$

(b) Fill in the gaps in the derivation of (44).

(c) For the case of electrons in a one-dimensional box of length L, perform the sum in (44) and obtain an explicit expression for the density-density correlator $C(x)$. What is the wavelength of the Friedel oscillations?

5. *Thermodynamics of the Fermi gas.* Using methods similar to those in the black-body radiation question — but remembering that you're now dealing with fermions — prove that the specific heat of the Fermi gas is given by

$$C_V = \gamma T \tag{51}$$

at temperatures $T \ll \varepsilon_F$, and obtain an expression for the Sommerfeld coefficient γ. (*Hint:* You will need to use an approximate form of the Fermi function or its derivative at some point. Proceed as far as possible before this approximation is made. You may like to try assuming that

$$\frac{1}{\cosh^2 x} \approx \begin{cases} 0 & |x| > 1, \\ & \\ 1 & |x| < 1. \end{cases} \quad)$$

2 Electron-electron interactions: Landau Fermi liquid theory

2.1 In two sentences...

A rough estimate of the relative magnitudes of the Coulomb and kinetic energies for electrons in a solid shows that the Coulomb energy is always dominant, and therefore the previous lecture's approach (based on non-interacting electrons) ought to fail radically. The fact that it does not is explained by Landau's Fermi liquid theory, which clarifies that the objects described in the above theory are not actually electrons, but *quasiparticles*.

2.2 Coulomb beats kinetic: a back-of-the-envelope calculation

In the previous lecture, we outlined second quantisation, and briefly discussed the Fermi gas theory of metals. The natural next step would be to reintroduce the Coulomb interaction between the electrons, and consider its effect. This would mean dealing not with a Fermi gas, but with a so-called **Fermi liquid**.

A simple way to estimate the importance of the inter-electron Coulomb interaction is to take its ratio with the kinetic energy of the electrons. A relevant parameter is the electron density in the material, N/V; we shall discuss this in terms of r_s, the radius of the sphere of space occupied by one electron (neglecting packing issues). Thus

$$\frac{4}{3}\pi r_s^3 = \frac{V}{N} \qquad \rightarrow \qquad r_s = \left(\frac{3}{4\pi}\frac{V}{N}\right)^{1/3}. \tag{52}$$

We shall approximate the kinetic energy as the energy that the electron would have if it were confined in this sphere,

$$E_K \approx \frac{\hbar^2}{8mr_s^2}, \tag{53}$$

and the Coulomb energy as the Coulomb repulsion of two electrons at the centre of neighbouring spheres, i.e. separated by a distance of roughly $2r_s$,

$$E_C \approx \frac{e^2}{8\pi\varepsilon_0 r_s}. \tag{54}$$

The ratio of these two is thus given by

$$\frac{E_C}{E_K} \approx \frac{8me^2}{8\pi\varepsilon_0\hbar^2} r_s = \frac{4r_s}{a_0}, \tag{55}$$

where a_0 is the familiar Bohr radius of atomic physics[18].

But in real materials r_s is typically a couple of Ångstroms, while a_0 is half of one. Therefore the ratio will almost always be significantly greater than one, i.e. the Coulomb energy will be larger than the kinetic energy[19]. This would seem to suggest that the behaviour of real metals should be dramatically different from that predicted by the Fermi-gas model; but experimentally that model actually does very well. For example, the linear electronic specific heat predicted by the Fermi-gas model ($C_V = \gamma T$: see exercises for previous lecture) is observed in a wide range of conductors. In this lecture, we shall discuss **Landau's Fermi liquid theory**, which provides insight into why the excitations of the strongly interacting Fermi liquid are — or at least can be — so much like those of the Fermi gas.

2.3 A perturbative approach; phase space restrictions

To begin with, let's play dumb and introduce the Coulomb interactions perturbatively, starting from the non-interacting Fermi gas. To do this, we modify the Fermi-gas Hamiltonian (32) by adding a term describing the effects of Coulomb interactions. Let us work out what this term should look like in second-quantised language.

The Coulomb term in the Hamiltonian may be written in terms of the electron field operators as[20]

$$H_C = \sum_{\sigma\tau} \int d\mathbf{r} \int d\mathbf{r}' \psi^\dagger_\sigma(\mathbf{r}')\psi^\dagger_\tau(\mathbf{r}) \frac{e^2}{4\pi\varepsilon_0|\mathbf{r}-\mathbf{r}'|} \psi_\tau(\mathbf{r})\psi_\sigma(\mathbf{r}'). \tag{56}$$

Inserting the decomposition of the electron field into momentum eigenstates (30), we find that

$$H_C = \sum_{\mathbf{qkp}\sigma\tau} V_{\mathbf{qkp}} c^\dagger_{\mathbf{q}\sigma} c^\dagger_{\mathbf{k}\tau} c_{\mathbf{p}\tau} c_{\mathbf{q}+\mathbf{k}-\mathbf{p},\sigma}. \tag{57}$$

You can calculate the actual matrix element $V_{\mathbf{qkp}}$ in the Exercises. Thus the full Hamiltonian for the Fermi liquid is

$$H = \underbrace{\sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma}}_{H_0} + \underbrace{\sum_{\mathbf{qkp}\sigma\tau} V_{\mathbf{qkp}} c^\dagger_{\mathbf{q}\sigma} c^\dagger_{\mathbf{k}\tau} c_{\mathbf{p}\tau} c_{\mathbf{q}+\mathbf{k}-\mathbf{p},\sigma}}_{H_C}, \tag{58}$$

with $\varepsilon_{\mathbf{k}}$ the free-electron dispersion relation used in the previous lecture. The H_C term has the effect of annihilating two electrons in momentum space, and recreating them at two other momenta, such that the total momentum is conserved in the process.

Our approach will be to calculate the effect of H_C on a low-lying excitation of the Fermi gas. This low-lying excitation consists of an electron sitting just above the Fermi surface[21], at

[18] What else could it have been, after all? We're making a constant with the dimensions of length to compare r_s to; but you can only make one constant with length dimensions from the Coulomb interaction, Planck's constant, and the electron mass, and that's a_0.

[19] The situation becomes even worse when we reflect that E_C is likely to be an *underestimate* of the Coulomb energy (since it ignores all electrons but the nearest) while E_K is likely to be an *overestimate* of the kinetic energy (since the electron is not exactly confined to the r_s-sphere)!

[20] One might think that, since the Coulomb interaction is a density-density interaction, the form of the Hamiltonian would be $\psi^\dagger\psi\psi^\dagger\psi$, rather than $\psi^\dagger\psi^\dagger\psi\psi$. But using the obvious density-density form leads to a non-zero contribution even when there is only one electron in the system, since the different 'parts' of the electron interact with each other. To prevent this, all the annihilation operators are moved to the right, and then the term will act only when at least two electrons are present. This is called **normal ordering**.

[21] Of course this electron must have a partner hole somewhere, since the total number of electrons should be conserved. But we shall assume that this hole is in some other part of $\mathbf{k}$-space, and does not participate in the scattering event to be described.

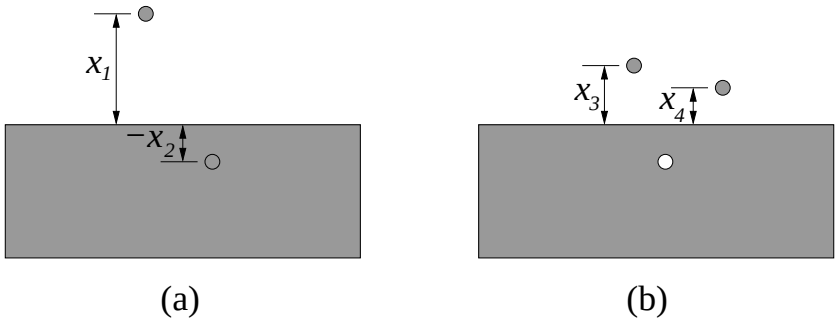


Figure 2: A pictorial representation of a Coulomb scattering event. (a) A low-lying excitation: an electron at energy x_1 above the Fermi level. Also indicated is the other electron that will be involved in the scattering. (b) The final state: two electrons at energies x_3 and x_4 above the Fermi level respectively, and a hole at energy x_2, where $x_2 \leqslant 0$, i.e. at energy $|x_2|$ below the Fermi level.

some energy $E = \varepsilon_F + x_1$. After a single action of H_C, this electron and one from inside the Fermi surface (at energy $E = \varepsilon_F + x_2$, where $x_2 \leqslant 0$) have been annihilated, and replaced by two new electrons at energies $E = \varepsilon_F + x_3$ and $E = \varepsilon_F + x_4$. Conservation of energy means that $x_1 + x_2 = x_3 + x_4$. This scattering event is shown in Fig. 2.

Let us calculate the rate at which such scatterings occur by applying the Fermi golden rule:

$$T_{i\to f} = \frac{2\pi}{\hbar} |\underbrace{\langle i| H_C |f\rangle}_{\text{matrix element}}|^2 \rho, \tag{59}$$

where ρ is the density of final states. We shall take the scattering matrix element $V_{\mathbf{qkp}}$ to be a constant, V_0; thus, the only relevant factor is the number of final states available. This is obtained by integrating over all possible values of x_2, x_3, and x_4 subject to the following constraints: (i) energy conservation; (ii) $x_2 \leqslant 0$; (iii) $x_3, x_4 \geqslant 0$. Thus the total number of final states is given by

$$N_f(x_1) = \nu^2 \int_{-\infty}^{0} dx_2 \int_0^{\infty} dx_3 \int_0^{\infty} dx_4 \, \delta(x_3 + x_4 - x_1 - x_2), \tag{60}$$

where ν is the density of states at the Fermi level. Performing the integrals, one finds that[22]

$$\begin{aligned} N_f(x_1) &= \nu^2 \int_{-\infty}^{0} dx_2 \int_0^{\infty} dx_3 \, \Theta(x_1 + x_2 - x_3) \\ &= \nu^2 \int_{-\infty}^{0} dx_2 \int_0^{x_1+x_2} dx_3 \, \Theta(x_1 + x_2) \\ &= \nu^2 \int_{-\infty}^{0} dx_2 \, (x_1 + x_2) \, \Theta(x_1 + x_2) \end{aligned}$$

[22]The function $\Theta(x)$ is now the usual step function, defined as 1 if $x > 0$, $\frac{1}{2}$ if $x = 0$, and 0 if $x < 0$.

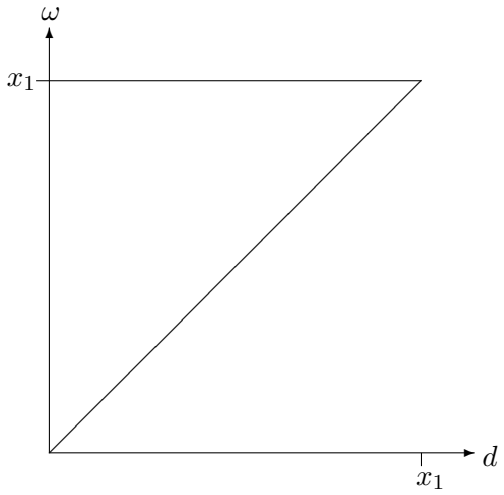


Figure 3: A pictorial argument for why the rate for electron scattering is proportional to the square of its energy above the Fermi level.

$$= \nu^2 \int_{-x_1}^{0} (x_1 + x_2) dx_2$$
$$= \frac{1}{2}\nu^2 x_1^2. \tag{61}$$

That derivation was rather mathematical, but there is a very simple graphical method that gives the same answer. Suppose we want to work out the number of ways that an electron with energy x_1 can scatter from the Fermi sea as shown in Fig. 2. We have to choose two things: (i) d, the depth within the Fermi sea from which we take the other electron, and (ii) ω, how much energy we give it. Now, the deepest we can go is x_1, because otherwise we can't excite the electron at all. Furthermore, once we've picked the depth d, we must give it at an energy $\omega \geqslant d$ to make sure it ends up outside the Fermi sea. These two constraints mark out a region in d-ω space shown in Fig. 3. The area of this triangle of permissible choices is clearly proportional to x_1^2; we just need to add a couple of density-of-states factors to get the dimensions right, yielding (61).

This is a highly significant result: let's pause for a moment to think about what it means. Firstly, since the Golden Rule scattering rate is proportional to $N_f(x_1)$, it means that the scattering rate vanishes as $x_1 \to 0$, i.e. as the excitation gets closer and closer to the Fermi surface. Furthermore, it vanishes as x_1^2. The excitation's energy, on the other hand, vanishes like x_1, which means that the ratio of scattering rate to energy is

$$\frac{\Gamma}{E} \sim x_1 \to 0 \qquad \text{as } x_1 \to 0. \tag{62}$$

Here is a heuristic argument for why the ratio (62) is the relevant one to consider: If we think of the excitation as a wave, its period is x_1^{-1}, whereas its damping time is x_1^{-2}, and therefore excitations near the Fermi surface perform many complete oscillation cycles before they get destroyed by H_C-scattering. In terms of oscillators, the ratio (62) is therefore Q^{-1}, where Q is the quality factor of the oscillator. Thus, the non-interacting picture becomes better and better

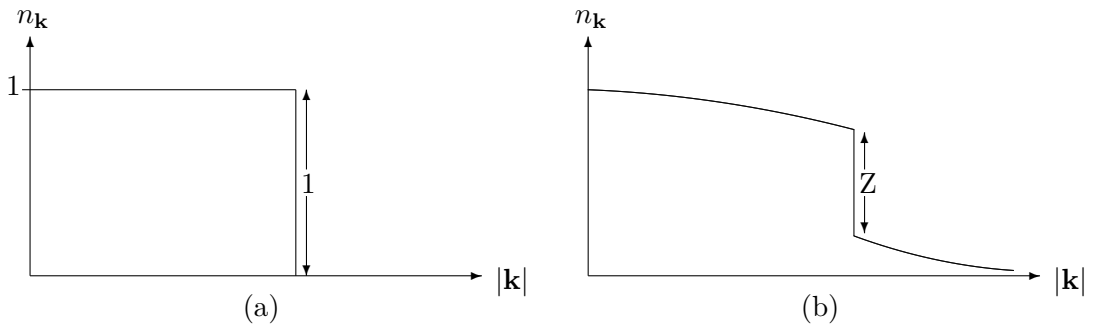


Figure 4: (a) The electron distribution $n_{\mathbf{k}}$ as a function of $|\mathbf{k}|$ at $T = 0$ in the Fermi gas case. (b) The electron distribution function in the presence of interactions. If a discontinuity at the Fermi wavevector survives, its height is the quasiparticle weight Z.

(i.e. the Q-factor increases without limit) as the excitation gets closer and closer to the Fermi surface. Experimentally, this usually corresponds to dealing with low temperatures.

2.4 Adiabatic continuity; energy functional and Landau parameters

The above argument is not enough to show that the Fermi surface is stable to such Coulomb interactions. This is because it *assumes* that the Fermi surface is perfect, separating a completely occupied region of **k**-space from a completely unoccupied one. But after the Coulomb interaction has been allowed to work for some time, this will no longer be the case. Excitations of the type we have just discussed will disrupt the upper part of the Fermi sea, leading to a distribution function for the electrons that looks more like that shown in Fig. 4(b).

The key idea of Landau's Fermi liquid theory is that, although this is true in the electron basis, there remains an alternative description of the interacting problem in which the distribution function still looks like Fig. 4(a), at least near the Fermi surface. This is achieved by thinking of it not as the distribution function of the original electrons, but as the distribution function of the true elementary excitations of the system, or **quasiparticles**. In this section, we shall expand on that argument, and introduce the Landau parameters that describe interactions between these quasiparticles; in the next, we shall begin to explore what happens when these interactions become strong enough to provoke a phase transition.

The assumption on which Landau Fermi liquid theory is based is that of **adiabatic continuity**. What this means is that, if we imagine turning up the strength of the Coulomb interactions from zero, the low-energy excitations of the Fermi liquid continue to be electron-like excitations labelled by a momentum **k** and a spin σ. These excitations we call **quasiparticles**. (They are not the only kinds of excitation in a Fermi liquid, since obviously a liquid can also support collective excitations such as sound modes that are not single-electron-like. The quasiparticle excitations, however, usually dominate the response of the system at low temperatures.) The way Landau theory works is to describe the deviation from the ground state energy, δE, in terms of the deviation from the ground state occupation function of these quasiparticles[23], $\delta n_{\mathbf{k}\sigma}$. If

[23]Notice — and this is important — that $\delta n_{\mathbf{k}\sigma}$ is the deviation from the ground-state distribution of *quasiparticles*, not of electrons. The ground-state distribution of quasiparticles is a step function by definition; the ground state distribution of electrons is usually more complicated, as shown in Fig. 4(b).

the deviation $\delta n_{\mathbf{k}\sigma}$ is small, we should be able to expand δE in powers of $\delta n_{\mathbf{k}\sigma}$:

$$\delta E\left(\{\delta n_{\mathbf{k}\sigma}\}\right) = \sum_{\mathbf{k}\sigma} A_{\mathbf{k}\sigma}\delta n_{\mathbf{k}\sigma} + \sum_{\mathbf{kq}\sigma\tau} F_{\mathbf{kq}\sigma\tau}\delta n_{\mathbf{k}\sigma}\delta n_{\mathbf{q}\tau} + \dots \tag{63}$$

The A coefficient is related to the change in the energy if we change the occupation of a single quasiparticle mode: thus we interpret it as the energy of that mode, $\varepsilon_{\mathbf{k}\sigma}$. The F coefficient is related to the effect of the Coulomb interactions, and in general we cannot obtain its analytic form. However, we can make a useful decomposition of its structure, as we shall now discuss.

In an isotropic Fermi liquid, there is no privileged direction, and so $F_{\mathbf{kq}\sigma\tau}$ can depend only on the relative angle between $\mathbf{k}$ and $\mathbf{q}$. Furthermore, at low energies, the magnitudes of $\mathbf{k}$ and $\mathbf{q}$ are expected to be very close to the Fermi wavevector k_F. Thus we can approximate

$$|\mathbf{k}| = |\mathbf{q}| = k_F; \qquad \mathbf{k}\cdot\mathbf{q} = k_F^2\cos\theta, \tag{64}$$

and expand the F-coefficient as a sum of Legendre polynomials:

$$F_{\mathbf{kq}\sigma\tau} = \sum_{l=0}^{\infty} F_{l\sigma\tau}P_l(\cos\theta). \tag{65}$$

Thus we can rewrite (63) in the form

$$\delta E\left(\{\delta n_{\mathbf{k}\sigma}\}\right) = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}\sigma}\delta n_{\mathbf{k}\sigma} + \sum_{\mathbf{kq}l\sigma\tau} F_{l\sigma\tau}P_l(\cos\theta)\delta n_{\mathbf{k}\sigma}\delta n_{\mathbf{q}\tau} + \dots \tag{66}$$

where θ is the angle between $\mathbf{k}$ and $\mathbf{q}$. Furthermore, in the absence of a magnetic field the interaction should depend only on the *relative* orientation of the quasiparticles' spins, and thus

$$\begin{aligned} F_{l\uparrow\uparrow} &= F_{l\downarrow\downarrow} \equiv F_l^s + F_l^a, \\ F_{l\downarrow\uparrow} &= F_{l\uparrow\downarrow} \equiv F_l^s - F_l^a, \end{aligned} \tag{67}$$

where the s and a superscripts stand for 'symmetric' and 'antisymmetric' respectively[24].

The $F_l^{(s/a)}$ are called the **Landau parameters**; they characterise the interaction between quasiparticles in the Fermi liquid. There are two more Landau parameters to add to this set: the **effective mass** and the **quasiparticle weight**.

- The effective mass is defined in terms of the group velocity of the quasiparticle excitations[25]:

  $$\mathbf{v}_{\mathbf{k}} = \frac{1}{\hbar}\nabla_{\mathbf{k}}\varepsilon_{\mathbf{k}} \equiv \frac{\hbar\mathbf{k}}{m^*}, \tag{68}$$

  with m^* the effective mass. In typical materials m^* is a few times the bare electron mass m; though there are some materials where $m^* \sim 10^3 m$, which are therefore called the 'heavy fermion materials'. These will be discussed briefly in lecture six.

[24] If the scattering is completely independent of the relative direction of the spins, then $F^a = 0$; if the up-down scattering amplitude is equal and opposite to the up-up scattering amplitude, then $F^s = 0$.

[25] Of course, this will not be the same as the group velocity for a free-electron wavepacket. Physically, one may think that as the electron moves it excites electron-hole pairs out of the Fermi sea, which increases its effective mass; the object made of the original electron plus this cloud of electron-hole pairs is the quasiparticle, and it will generally move more slowly than its free-electron counterpart.

- The quasi-particle weight, Z, is a different beast: rather than measuring the nature of interactions among the quasiparticles, it measures the difference between what the electrons are doing and what the quasiparticles are doing. Clearly this is some measure of the extent to which the electrons are also participating in non-quasiparticle excitations such as sound waves. A quasiparticle weight $Z = 1$ corresponds to the Fermi gas (all electrons doing what all quasiparticles are doing $\rightarrow$ electrons and quasiparticles identical), whereas a quasiparticle weight $Z \rightarrow 0$ signals the imminent breakdown of the Landau Fermi liquid picture. A simple measure of Z is to examine the electron distribution function at $T = 0$. In the non-interacting case, the step at the Fermi surface has height 1; in the interacting case, this step (called the **Migdal discontinuity** if it still exists) has height Z. This is shown in Fig. 4.

2.5 Instabilities of the Fermi liquid: Stoner ferromagnetism

Of course, not all materials in nature are paramagnetic metals; not even all crystalline materials are. There must therefore be ways in which the Fermi liquid can break down and undergo a transition into another state. Fortunately, like Cassandra, Landau's Fermi liquid theory has the power to predict its own death: we shall discuss in this section how this works. The essential idea is the same that is used in assessing the stability of any mechanical system. Namely, we try out the kind of distortion we have in mind: if the energy goes up, the system is stable with respect to that kind of distortion; if the energy goes down, it is unstable. As an example, let's consider the possibility of a **ferromagnetic** instability, i.e. the spontaneous uniform magnetisation of the system without breaking any spatial symmetries.

In order to magnetise, the system of fermions must increase the number of up spins compared to the number of down spins, while keeping the total conserved. Therefore, the Fermi surface must split into a majority-spin Fermi surface (whose Fermi energy increases slightly) and a minority-spin Fermi surface (whose Fermi energy decreases slightly). The $\delta n_{\mathbf{k}\sigma}$ that describes this behaviour is

$$\delta n_{\mathbf{k}\sigma} = \begin{cases} 1 & \sigma = \uparrow,\ k_F < |\mathbf{k}| \leqslant k_F + \alpha; \\ -1 & \sigma = \downarrow,\ k_F - \alpha < |\mathbf{k}| \leqslant k_F; \\ 0 & \text{otherwise.} \end{cases} \tag{69}$$

To calculate the energy associated with such a change, we must substitute (69) into (63). Before we do so, let us think about what the result will be.

The single-quasiparticle term in (63) must be positive, since we are disturbing the distribution from the original filled Fermi sea, which by definition minimises the single-quasiparticle energy. Therefore, if an instability is to occur, some of the interactions in the second term must make a negative contribution to the energy large enough to cancel out the single-quasiparticle cost of the distortion. Thus, there is presumably a threshold value of some of the Landau parameters beyond which spontaneous magnetisation takes place. The question is, which Landau parameters are those, and what is the numerical value of that threshold?

To answer that question, we substitute (69) into (63); note that, since $\delta n_{\mathbf{k}\sigma}$ is spherically symmetric in $\mathbf{k}$, the only Legendre polynomial that does not vanish when the $\mathbf{k}$- and $\mathbf{q}$-sums are carried out is the $l = 0$ one:

$$\delta E = \sum_{\mathbf{k}_+} \varepsilon_{\mathbf{k}} - \sum_{\mathbf{k}_-} \varepsilon_{\mathbf{k}} + \sum_{\mathbf{k}_-} \sum_{\mathbf{q}_-} F_{0\downarrow\downarrow} + \sum_{\mathbf{k}_+} \sum_{\mathbf{q}_+} F_{0\uparrow\uparrow} - \sum_{\mathbf{k}_-} \sum_{\mathbf{q}_+} F_{0\downarrow\uparrow} - \sum_{\mathbf{k}_+} \sum_{\mathbf{q}_-} F_{0\uparrow\downarrow}, \tag{70}$$

where a $\pm$ subscript on a summation variable means that it runs over a thin shell from k_F to $k_F \pm \alpha$. Each summation therefore introduces a factor α, which means that the leading terms

in δE will be of order α^2. Replacing the sums by integrals in the usual way, and noting that the volume of momentum space occupied by a single state is $8\pi^3/V$, we thus obtain

$$\begin{aligned}
\delta E &= \frac{k_F^2 V\alpha}{\pi^2}\varepsilon_{k+\frac{\alpha}{2}} - \frac{k_F^2 V\alpha}{\pi^2}\varepsilon_{k-\frac{\alpha}{2}} + \left(\frac{k_F^2 V\alpha}{2\pi^2}\right)^2 \left[F_{0\uparrow\uparrow} + F_{0\downarrow\downarrow} - F_{0\uparrow\downarrow} - F_{0\downarrow\uparrow}\right] \\
&= \frac{k_F^2 V\alpha^2}{\pi^2}\left.\frac{\partial\varepsilon_k}{\partial k}\right|_{k=k_F} + \left(\frac{k_F^2 V\alpha}{\pi^2}\right)^2 F_0^a \\
&= \frac{k_F^2 V\alpha^2}{\pi^2}\frac{\hbar^2 k_F}{m^*} + \left(\frac{k_F^2 V\alpha}{\pi^2}\right)^2 F_0^a, \qquad (71)
\end{aligned}$$

where we have used the definition of the effective mass (68) and the definition of the Landau parameters (67). Recalling that the density of states at the Fermi surface, ν, is given by

$$\nu = \frac{k_F V m^*}{\pi^2\hbar^2}, \qquad (72)$$

we discover that

$$\delta E = \frac{k_F^3 V\hbar^2\alpha^2}{\pi^2 m^*}\left(1 + \nu F_0^a\right). \qquad (73)$$

For this to be an instability, we require that $\delta E < 0$, which happens when

$$\nu F_0^a < -1. \qquad (74)$$

This is called the **Stoner criterion**, and this method of a system's magnetising is called the **Stoner mechanism**. Notice, however, that this low-order theory of small distortions of the Fermi liquid does *not* permit us to describe the state into which the system goes as a result of the instability: all it can do is signal that the isotropic, unmagnetised Fermi liquid has become unstable. There are various other distortions of the Fermi surface to which the liquid may succumb; for another example, see the Exercises.

2.6 Relations between the Landau parameters

An important question to ask about Landau's Fermi liquid theory is whether it is just a convenient way of summarising the effects of interactions, or actually has predictive power. Equation (74), for example, provides a criterion for a ferromagnetic instability; but without knowing the value of F_0^a, we can never predict whether or not it will happen in a given metallic system. Landau's theory does not offer microscopic calculations of the Landau parameters; rather, it was developed as an alternative to them. But is it actually offering us anything?

The answer is yes, and it relies on the observation that there are more observables than there are Landau parameters. In other words, one can prove certain relationships that ought to hold between the Landau parameters, and then measure the parameters (using independent experiments) to check whether they do. An instance of this is the following relation between the effective mass and the Landau parameter F_1^s:

$$\frac{m^*}{m} = 1 + \frac{\nu F_1^s}{3}. \qquad (75)$$

Such relations are usually proved by using some global symmetry of the system; this one, for example, follows from Galilean invariance. We shall merely outline the scheme of the proof here.

The 'distortion' of the Fermi surface we consider is a uniform shift of the centre of the sphere in the k_x direction, by an amount $\mathbf{k} = k\hat{\mathbf{x}}$. This is shown in Fig. 5. On the one hand, the energy

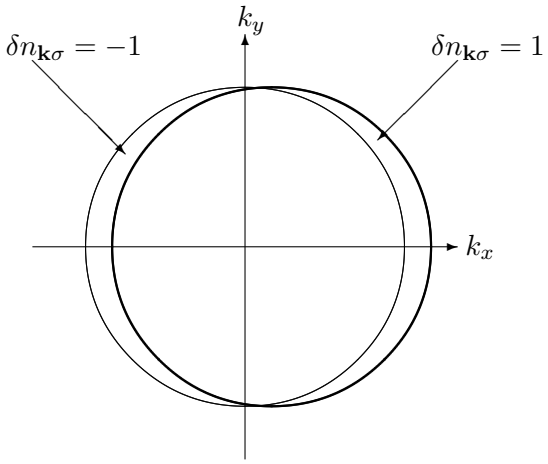


Figure 5: The 'distortion' of the Fermi surface used in proving (75): a uniform displacement of the Fermi sphere in momentum space. The thin line represents the original position of the Fermi surface, the thick line its new position. The 'lemon-slices' at either side are the regions of **k**-space where the population has changed.

of this distortion (by Galilean invariance) must be given by

$$\delta E = N \frac{\hbar^2 k^2}{2m}, \tag{76}$$

since it constitutes a Galilean boost of all the electrons[26] into a frame moving with momentum $\mathbf{p} = \hbar\mathbf{k}$. On the other hand, it is a distortion whose energy can be calculated using the Landau energy functional (63). Since the symmetry of the distortion is $l = 1$, this is the only Landau parameter that appears in the answer (except the effective mass); furthermore, since the distortion is independent of spin, the parameter F_1^s is the relevant one. Deriving an explicit expression for the 'distortion' $\delta n_{\mathbf{k}\sigma}$, and substituting it into (63), one finds

$$\delta E = N \left(\frac{\hbar^2 k^2}{2m^*} + \frac{k_F V}{2\pi^2} k^2 \frac{F_1^s}{3} \right). \tag{77}$$

Comparing (76) and (77), one finds the relation (75).

2.7 Exercises

1. *Phase space restrictions.*

   (a) Calculate the matrix element $V_{\mathbf{qkp}}$ from (58). Hence critically assess our assumption that V is approximately constant for processes taking place near the Fermi surface.

   (b) Does the electron-electron scattering rate near the Fermi surface depend on the number of spatial dimensions? If so, how?

[26] Note, the electrons: *not* the quasiparticles.

2. *Pomeranchuk instabilities.* A **Pomeranchuk instability** is defined as the instability of a Fermi liquid to a spontaneous distortion in the shape of its Fermi surface. (This is distinct from the Stoner instability, in which the Fermi surface for each spin species remains spherical but changes in volume.) Consider, for example, a distortion where the Fermi wavevector changes with polar angle in **k**-space as follows:

$$k_F(\theta) = k_{F0} + \alpha P_2(\cos\theta) - \frac{\alpha^2}{5k_{F0}}, \tag{78}$$

where k_{F0} is the original Fermi wavevector, θ is the polar angle in **k**-space, $P_2(\cos\theta) = \frac{3}{2}\cos^2\theta - \frac{1}{2}$ is the $l = 2$ Legendre polynomial, α is a small parameter, and the last term is necessary to ensure conservation of the total number of quasiparticles.

   (a) Write down an explicit expression for the $\delta n_{\mathbf{k}\sigma}$ corresponding to the distortion (78), assuming that the distortion is the same for both spin species.

   (b) Make a sketch of the distorted Fermi surface.

   (c) Using the Landau energy functional (63), calculate the energy cost of such a distortion up to and including terms of order α^2.

   (d) Hence obtain the instability criterion in terms of the Landau parameters $F_{l\sigma\tau}$. Comment on your answer in comparison with the Stoner criterion (74) derived above.

(I am indebted to Dr P.H. Conlon, formerly of the University of Oxford, for providing corrections to the original version of this problem and its solution.)

3 Magnetic impurities: the Anderson model and the Kondo effect

3.1 In two sentences...

In some metals, a Curie magnetic susceptibility (reminiscent of that of a localised spin) is observed, and is attributed to the presence of magnetic impurities. In this lecture we study how these impurities become magnetic (which seems naïvely as if it ought to be impossible), and study the formation of magnetic singlets between the impurities and the conduction sea at low temperatures: the so-called Kondo effect.

3.2 Magnetic susceptibilities: a quick recap

Let us begin this lecture, which will mainly concern magnetic effects, by considering the magnetic properties of two simple systems: the individual localised spin, and the Fermi gas.

- *Localised spin.* The Hamiltonian for a spin S in a magnetic field B is

$$H = -\mu_B B S^z, \tag{79}$$

where μ_B is the Bohr magneton, and without loss of generality we have chosen the field direction as our quantisation axis. We shall calculate the magnetic susceptibility for the simplest non-trivial case, i.e. $S = \frac{1}{2}$. In this case there are only two eigenvalues of S^z, given by $S^z = \pm\frac{1}{2}$, and so the partition function is simply

$$Z = \exp\left(\frac{\mu_B B}{2k_B T}\right) + \exp\left(-\frac{\mu_B B}{2k_B T}\right) = 2\cosh\left(\frac{\mu_B B}{2k_B T}\right). \tag{80}$$

The free energy is therefore

$$F = -k_B T \ln Z = -k_B T \ln 2 - k_B T \ln \cosh \left(\frac{\mu_B B}{2k_B T} \right); \tag{81}$$

the magnetisation is given by the B-derivative of F, i.e.

$$M = -\frac{\partial F}{\partial B} = \frac{\mu_B}{2} \tanh \left(\frac{\mu_B B}{2k_B T} \right). \tag{82}$$

Note that this expression saturates at the value $\frac{\mu_B}{2}$ for fields much greater than the temperature. On the other hand, for fields $\mu_B B \ll 2k_B T$, it becomes

$$M = \frac{\mu_B^2 B}{4k_B T}, \tag{83}$$

and hence the linear magnetic susceptibility is

$$\chi \equiv \left. \frac{\partial M}{\partial B} \right|_{B=0} = \frac{\mu_B^2}{4k_B T}. \tag{84}$$

This is the **Curie law**. Notice that the susceptibility diverges as $T \to 0$: this reflects the fact that, at zero temperature, the spin will be completely aligned by an arbitrarily small field.

- *Fermi gas.* To calculate the magnetisation of the Fermi gas at small fields, we consider the Hamiltonian in the presence of the appropriate Zeeman term[27]

$$H = \sum_{\mathbf{k}\sigma} \left(\varepsilon_{\mathbf{k}} - \sigma \mu_B B \right) c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma}, \tag{85}$$

where to interpret the notation one must take $\sigma = \pm \frac{1}{2}$. The magnetic field shifts the dispersion relations for the up- and down-spin electrons differently, and thus although they are filled to the same chemical potential μ, there are more of one species than of the other. This situation is sketched in Fig. 6. To be quantitative, at $T = 0$ the total number of up-spins is now

$$N_\uparrow = \int_{-\infty}^{\mu} \nu \left(\varepsilon + \frac{\mu_B B}{2} \right) d\varepsilon, \tag{86}$$

while the number of down-spins is

$$N_\downarrow = \int_{-\infty}^{\mu} \nu \left(\varepsilon - \frac{\mu_B B}{2} \right) d\varepsilon. \tag{87}$$

[27] Might there not, you ask, also be orbital effects of the magnetic field? After all, the electrons are charged and should be subject to a Lorentz force as well as a Zeeman splitting. Such orbital effects are indeed present, and are diamagnetic. In the isotropic system, the diamagnetic contribution due to orbital parts is exactly one third the size of the paramagnetic contribution due to the Zeeman effect — since they both come ultimately from the same term in the Dirac equation, such a simple proportionality should not be a surprise. However, the presence of the crystal lattice tends to quench the diamagnetism while leaving the Zeeman paramagnetism largely unaffected; hence we tend to neglect the diamagnetic orbital contribution in simple calculations such as these.

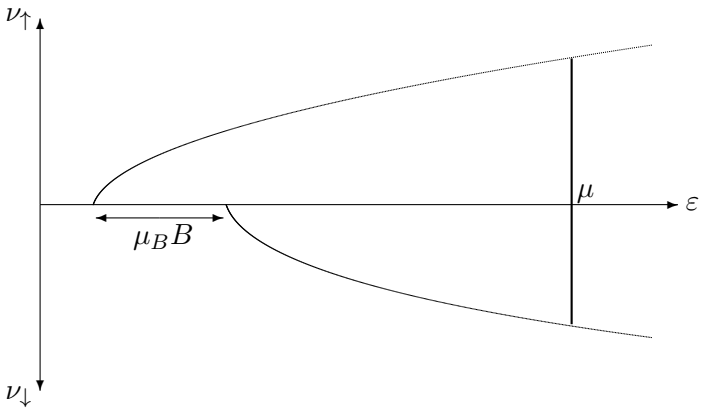


Figure 6: A sketch of the process whereby a Fermi gas magnetises in a field. The up-spin and down-spin bands shift in opposite directions, leaving an overall excess of up spins when the bands are filled to the same Fermi level μ. The horizontal axis is energy, while the two vertical axes are the densities of states for up- and down-spin electrons respectively.

Making a shift of variable in the integrals, we find that

$$\begin{aligned} N_\uparrow &= \int_{-\infty}^{\mu+\frac{\mu_B B}{2}} \nu(\varepsilon)d\varepsilon, \\ N_\downarrow &= \int_{-\infty}^{\mu-\frac{\mu_B B}{2}} \nu(\varepsilon)d\varepsilon, \end{aligned} \tag{88}$$

and hence

$$M = \frac{\mu_B}{2}(N_\uparrow - N_\downarrow) = \frac{\mu_B}{2}\int_{\mu-\frac{\mu_B B}{2}}^{\mu+\frac{\mu_B B}{2}} \nu(\varepsilon)d\varepsilon. \tag{89}$$

For sufficiently small fields, the integrand may be approximated by a constant, leading to

$$M \approx \frac{\mu_B^2 \nu B}{2}, \tag{90}$$

where ν is the density of states at the Fermi energy. Hence the linear susceptibility is given by

$$\chi = \frac{\mu_B^2 \nu}{2}. \tag{91}$$

This is known as the **Pauli susceptibility**. Although this simple argument was strictly for $T = 0$, it is not hard to show[28] that the result (91) remains valid provided that $k_B T \ll \mu$.

[28]To see how the proof works, think of what the diagram in Fig. 6 would look like at non-zero temperature. All that would happen is that the step function at the chemical potential would deform into a non-zero-T Fermi distribution function. But it would do so in the same way for both spin species, so the effect would cancel out when calculating the magnetisation. This argument fails when the bottom of the band starts to get depopulated, i.e. when $k_B T \sim \mu$.

Let us conclude this section with a brief interpretation of the above results. Why does a localised spin have a Curie susceptibility, while a Fermi sea has a Pauli susceptibility? The basic answer is that the Fermi sea has an intrinsic energy scale (the Fermi energy, ε_F, arising from the Pauli exclusion principle), whereas the single spin does not. Hence, at low temperatures it becomes infinitely easy to polarise the single spin, while there remains a resistance — again due to the exclusion principle — to polarising the Fermi sea. This becomes particularly clear if we recall that

$$\nu \sim \frac{N}{\varepsilon_F}; \tag{92}$$

we see, then, that the expression (90) is determining the magnetisation of the Fermi sea by comparing the applied field scale $\mu_B B$ to the Fermi energy ε_F. In the local spin case (83), the comparison is between the applied field $\mu_B B$ and the temperature scale $k_B T$, since the spin has no intrinsic energy scale of its own.

3.3 Why do magnetic impurities exist?: the Anderson model

It was observed early in the twentieth century that some metals, despite being expected to show the temperature-independent Pauli susceptibility (91), in addition exhibit Curie-like behaviour over a wide range of temperature. It is natural to suppose that, situated within the metal, are some impurity atoms that are behaving as localised spins, giving rise to a total susceptibility

$$\chi = \chi_0 + \frac{A}{T}, \tag{93}$$

with the two terms being due to the conduction sea and the magnetic impurities respectively. Even though the coefficient A is small (since it is proportional to the concentration of impurities), the second term becomes noticeable as the temperature is reduced because it is increasing while the first one is constant.

However, there is an apparent flaw with this explanation, as there is a simple argument that suggests that magnetic impurities should not be able to exist in metals. In fact, it's very simple: if there's an impurity level with energy ε_0, then either $\varepsilon_0 < \mu$ or $\varepsilon_0 > \mu$. In the former case, the energy is minimised by moving two electrons from the Fermi surface into the impurity level; in the latter case, the energy is minimised by leaving the impurity level empty. In either case, the configuration is non-magnetic: this is obvious where the impurity is empty; when it is doubly occupied, the antisymmetry of the wavefunction means that the magnetic state is

$$\frac{1}{\sqrt{2}} \left(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle \right), \tag{94}$$

which is a singlet ($S = 0$) state and so also non-magnetic. To put that argument in second-quantised language, the Hamiltonian

$$H = \underbrace{\sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma}}_{\text{conduction sea}} + \underbrace{\varepsilon_0 \sum_{\sigma} d^\dagger_\sigma d_\sigma}_{\text{impurity level}} + \underbrace{\sum_{\mathbf{k}\sigma} \left(V_{\mathbf{k}} c^\dagger_{\mathbf{k}\sigma} d_\sigma + V^*_{\mathbf{k}} d^\dagger_\sigma c_{\mathbf{k}\sigma} \right)}_{\text{tunnelling}} \tag{95}$$

never develops a Curie term in its susceptibility, which is easy to prove.

So how come the Curie term appears in measurements on some metals? The answer was given by Anderson in the 1950s, and it has to do with the Coulomb interaction between the electrons. In the conduction sea (which is usually made of electrons from s-orbitals), this

Coulomb interaction has little effect because the s-orbitals are fairly broad; but an impurity level is often a d- or an f-orbital. These have much more tightly compressed wavefunctions, and so the Coulomb cost of doubly occupying the level may be large. Thus, in addition to the Hamiltonian (95), there should be a term that raises the energy if an up- and a down-spin electron are found on the impurity together. The Hamiltonian then becomes

$$H = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} + \varepsilon_0 \sum_\sigma d^\dagger_\sigma d_\sigma + \sum_{\mathbf{k}\sigma} \left(V_{\mathbf{k}} c^\dagger_{\mathbf{k}\sigma} d_\sigma + V^*_{\mathbf{k}} d^\dagger_\sigma c_{\mathbf{k}\sigma} \right) + U d^\dagger_\uparrow d_\uparrow d^\dagger_\downarrow d_\downarrow, \tag{96}$$

where U is the extra energy cost of doubly occupying the impurity level. This model is called the **Anderson model**. As we shall see, the addition of the U-term transforms it from an easily solvable model into one with remarkably rich physics, even though it describes only a single impurity in the middle of a Fermi gas.

It is common to make the further simplification that the tunnelling matrix elements $V_{\mathbf{k}}$ are in fact (a) real, and (b) independent of $\mathbf{k}$, in which case the model becomes[29]

$$H = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} + \varepsilon_0 \sum_\sigma d^\dagger_\sigma d_\sigma + V \sum_{\mathbf{k}\sigma} \left(c^\dagger_{\mathbf{k}\sigma} d_\sigma + d^\dagger_\sigma c_{\mathbf{k}\sigma} \right) + U d^\dagger_\uparrow d_\uparrow d^\dagger_\downarrow d_\downarrow. \tag{97}$$

A simple argument to show that this model does what it's supposed to do, i.e. permits the existence of magnetic impurities in a metallic host, is developed in the Exercises.

3.4 The Kondo limit

A particularly clean treatment of the effect of a magnetic impurity can be made by ensuring that its charge never really fluctuates, i.e. by making the empty and doubly-occupied states so high in energy that they are essentially never visited. One is left with a two-state system on the impurity, where the states are (a) singly-occupied with spin down and (b) singly-occupied with spin up. This two-state system is clearly a quantum spin with $S = \frac{1}{2}$, and thus can be described by a spin operator $\mathbf{S} = (S^x, S^y, S^z)$. This limit is called the **Kondo limit**, and the resulting Hamiltonian is

$$H = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} + J \mathbf{s}_0 \cdot \mathbf{S}, \tag{98}$$

where $\mathbf{S}$ is the spin-$\frac{1}{2}$ operator representing the impurity spin, and[30]

$$\mathbf{s}_0 \equiv \sum_{\sigma\tau} \psi^\dagger_\sigma(\mathbf{r}=\mathbf{0}) \boldsymbol{\sigma}_{\sigma\tau} \psi_\tau(\mathbf{r}=\mathbf{0}) \propto \sum_{\mathbf{k}\mathbf{q}\sigma\tau} c^\dagger_{\mathbf{k}\sigma} \boldsymbol{\sigma}_{\sigma\tau} c_{\mathbf{q}\tau} \tag{99}$$

is the spin density of the conduction sea at the origin (where the impurity is). Substituting (99) into (98), we obtain the model

$$H = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} + \frac{J}{2} \sum_{\mathbf{k}\mathbf{q}\sigma\tau} c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{q}\tau} \mathbf{S} \cdot \boldsymbol{\sigma}_{\sigma\tau}. \tag{100}$$

For a derivation of an expression for J in terms of the Anderson model parameters, see the Exercises.

[29] This corresponds to the case of a point-like impurity located at the origin.

[30] The notation gets a little thick here. The entity $\boldsymbol{\sigma}$ is a vector of Pauli matrices, and *each* of these matrices has the indices σ and τ. Hence $\mathbf{S} \cdot \boldsymbol{\sigma}_{\sigma\tau} = S^x \sigma^x_{\sigma\tau} + S^y \sigma^y_{\sigma\tau} + S^z \sigma^z_{\sigma\tau} = S^z \sigma^z_{\sigma\tau} + \frac{1}{2}\left(S^+ \sigma^-_{\sigma\tau} + S^- \sigma^+_{\sigma\tau} \right)$. The final representation is most useful for practical calculations.

The obvious way to treat the model (100) is via perturbation theory in the coupling J, since the eigenstates of the $J = 0$ model are easy to describe. For example, we could calculate the extra resistance in the metallic host due to scattering from the impurity; this turns out to be[31]

$$R_{\rm imp} = R_0 J^2 \left[1 - 2\nu J \ln\left(\frac{k_B T}{\varepsilon_F}\right)\right] + O(J^4), \tag{101}$$

where ν is the conduction-electron density of states at the Fermi surface, and R_0 is a constant that can be calculated. The logarithm in the J^3 term is commonly known as the **Kondo logarithm**: it appears in calculations for impurities that have internal degrees of freedom, but not for those that don't. Disturbingly, the Kondo logarithm diverges as the temperature is lowered, and at some stage will clearly overpower the J^2 term, thus invalidating the perturbation expansion. All we can do is to calculate the temperature at which this failure occurs: we can estimate it roughly by asking at what temperature the J^2 and J^3 terms become equal. That happens when

$$1 = -2\nu J \ln\left(\frac{k_B T}{\varepsilon_F}\right) \qquad \rightarrow \qquad T = \frac{\varepsilon_F}{k_B} \exp\left(-\frac{1}{2\nu J}\right) \equiv T_K; \tag{102}$$

the temperature T_K is called the **Kondo temperature**, and in real metals tends to be a few kelvin, though is sometimes much higher. Notice that the Kondo temperature has an essential singularity at $J = 0$, and hence cannot be expanded in powers of J.

We should mention that this logarithmic increase of scattering as temperature is lowered can be seen in many materials, and was one of the original motivations for studying the Kondo model (100). Including a T^5 term due to electron-phonon scattering, the resistivity of a metal with magnetic impurities becomes

$$\rho(T) = AT^5 + c\rho_0 J^2 \left[1 - 2\nu J \ln\left(\frac{k_B T}{\varepsilon_F}\right)\right], \tag{103}$$

where c is the concentration of the impurities. This resistivity has a minimum at non-zero temperature:

$$\frac{\partial \rho}{\partial T} = 5AT^4 - \frac{2c\rho_0 \nu J^3}{T} = 0 \qquad \rightarrow \qquad T_{\rm min} \sim c^{1/5}. \tag{104}$$

This characteristic dependence of the temperature of the minimum on the impurity concentration fits well with experiments on dilute alloys, and was thus a strong piece of evidence in favour of Kondo's explanation of the resistance minimum.

3.5 Variational approach: the Varma-Yafet solution

But what happens at temperatures $T \ll T_K$? — what is the ground state of the problem? Since perturbation theory in J breaks down at temperatures below T_K, we need a non-perturbative technique to describe the low-temperature state. Varma and Yafet proposed a simple scheme: start from the Anderson model, and construct a variational state that explicitly forbids double occupation of the impurity level. Optimise this state by minimising its average energy, and then send $\varepsilon_0 \rightarrow -\infty$ to achieve the Kondo limit.

[31] Like most perturbative calculations, this one is tedious in its details, but not actually difficult. A full account is given in chapter 2 of A.C. Hewson, *The Kondo Problem to Heavy Fermions* (Cambridge University Press, 1993).

The state they proposed is given by[32]

$$|\psi\rangle = a_0 |FS\rangle + \sum_{\mathbf{k}} a_{\mathbf{k}} \left(d^\dagger_\uparrow c_{\mathbf{k}\uparrow} + d^\dagger_\downarrow c_{\mathbf{k}\downarrow} \right) |FS\rangle \,, \tag{105}$$

where the state $|FS\rangle$ consists of a filled Fermi sea and an empty impurity level. Note that the state $|\psi\rangle$ never has more than one electron on the impurity level, by construction. The expectation value of the energy in this state is[33]

$$E_V = \frac{\langle\psi| H_A |\psi\rangle}{\langle\psi|\psi\rangle}, \tag{106}$$

where H_A is the Anderson Hamiltonian (97). To calculate this, let us construct $H_A |\psi\rangle$, keeping only those terms that have either no holes or one hole in the Fermi sea — since all others will have no overlap with $|\psi\rangle$ and hence will not contribute to E_V. This restricted action of H_A on $|\psi\rangle$ yields:[34]

$$\begin{aligned} PH_A |\psi\rangle &= PH_A \left[a_0 |FS\rangle + \sum_{\mathbf{k}} a_{\mathbf{k}} \left(d^\dagger_\uparrow c_{\mathbf{k}\uparrow} + d^\dagger_\downarrow c_{\mathbf{k}\downarrow} \right) |FS\rangle \right] \\ &= a_0 E_0 |FS\rangle + \sum_{\mathbf{k}} a_{\mathbf{k}} \left(E_0 - \varepsilon_{\mathbf{k}} + \varepsilon_0 \right) \left(d^\dagger_\uparrow c_{\mathbf{k}\uparrow} + d^\dagger_\downarrow c_{\mathbf{k}\downarrow} \right) |FS\rangle \\ &\quad + a_0 V \sum_{\mathbf{k}} \left(d^\dagger_\uparrow c_{\mathbf{k}\uparrow} + d^\dagger_\downarrow c_{\mathbf{k}\downarrow} \right) |FS\rangle + 2V \sum_{\mathbf{k}} a_k |FS\rangle \,, \end{aligned} \tag{107}$$

and hence the variational energy E_V is given by[35]

$$E_V = \frac{a_0^2 E_0 + 4 a_0 V \sum_{\mathbf{k}} a_{\mathbf{k}} + 2 \sum_{\mathbf{k}} a_{\mathbf{k}}^2 \left(E_0 - \varepsilon_{\mathbf{k}} + \varepsilon_0 \right)}{a_0^2 + 2 \sum_{\mathbf{k}} a_{\mathbf{k}}^2}. \tag{108}$$

Minimising this with respect to a_0 and the $a_{\mathbf{k}}$, we obtain after some manipulation the set of coupled equations:

$$a_0 V + a_{\mathbf{k}} \left(E_0 - \varepsilon_{\mathbf{k}} + \varepsilon_0 - E_V \right) = 0; \tag{109}$$

$$a_0 \left(E_0 - E_V \right) + 2V \sum_{\mathbf{k}} a_{\mathbf{k}} = 0. \tag{110}$$

The quantity of physical interest is the binding energy between the impurity and the conduction sea, which is given by

$$\varepsilon \equiv E_V - E_0 - \varepsilon_0; \tag{111}$$

substituting this into (109) and (110), we can eliminate E_0:

$$a_0 V + a_{\mathbf{k}} \left(-\varepsilon - \varepsilon_{\mathbf{k}} \right) = 0; \tag{112}$$

$$a_0 \left(-\varepsilon_0 - \varepsilon \right) + 2V \sum_{\mathbf{k}} a_{\mathbf{k}} = 0. \tag{113}$$

[32] Note that, in this section, a_0 and all of the $a_{\mathbf{k}}$ are just numerical coefficients, not bosonic operators.

[33] Note the normalisation factor in the denominator. Since we will be varying the a_0 and $a_{\mathbf{k}}$ independently, this factor is needed to avoid the mathematics' 'minimising' the energy by simply making all the coefficients infinitely large.

[34] For mathematical correctness, we have put a projector in front to represent the fact that we keep only those states in the zero-and-one-hole manifold.

[35] We have taken a_0 and all the $a_{\mathbf{k}}$ to be real.

Equation (112) gives us

$$a_{\mathbf{k}} = \frac{a_0 V}{\varepsilon + \varepsilon_{\mathbf{k}}}; \tag{114}$$

substituting this into (113) yields

$$-a_0 (\varepsilon + \varepsilon_0) + 2a_0 V^2 \sum_{\mathbf{k}} \frac{1}{\varepsilon + \varepsilon_{\mathbf{k}}} = 0. \tag{115}$$

Assuming that $a_0 \neq 0$, we obtain an implicit equation for the bound state energy:

$$\varepsilon = -\varepsilon_0 + 2V^2 \sum_{\mathbf{k}} \frac{1}{\varepsilon_{\mathbf{k}} + \varepsilon}. \tag{116}$$

As usual, we can transform the $\mathbf{k}$-sum into an integral; noting that it includes only occupied states of the Fermi sea, we then obtain

$$\varepsilon = -\varepsilon_0 + 2V^2 \int_{-\infty}^{0} \nu(\chi) \frac{1}{\varepsilon + \chi} d\chi, \tag{117}$$

where $\nu(\chi)$ is the conduction-electron density of states and we have adopted the Fermi level as our zero of energy. For simplicity, let $\nu(\chi)$ take the form

$$\nu(\chi) = \begin{cases} \nu & \chi \geqslant -\varepsilon_F; \\ 0 & \text{otherwise,} \end{cases} \tag{118}$$

since on this energy scale $-\varepsilon_F$ is the bottom of the Fermi sea. Then we can do the integral in (117):

$$\begin{aligned} \varepsilon &= -\varepsilon_0 + 2V^2 \nu \int_{-\varepsilon_F}^{0} \frac{1}{\varepsilon + \chi} d\chi \\ &= -\varepsilon_0 + 2V^2 \nu \Big[\ln |\varepsilon + \chi| \Big]_{-\varepsilon_F}^{0} \\ &= -\varepsilon_0 + 2V^2 \nu \ln \left| \frac{\varepsilon}{\varepsilon - \varepsilon_F} \right| \\ &\approx -\varepsilon_0 - 2V^2 \nu \ln \left(\frac{\varepsilon_F}{|\varepsilon|} \right), \end{aligned} \tag{119}$$

where in the final step we have made the (usually very reliable) approximation $\varepsilon_F \gg |\varepsilon|$. The result (119) is, of course, the Kondo logarithm again, but this time there is no divergence.

Let us find out what this binding energy is in the Kondo limit, i.e. when $\varepsilon_0 \to -\infty$. In this case, the binding energy must become rather small, since the logarithm must compensate the $-\varepsilon_0$ term on the right-hand side of (119); in that case, we can neglect the ε on the left-hand side, to find the approximate solution

$$\varepsilon_0 = -2V^2 \nu \ln \left(\frac{\varepsilon_F}{|\varepsilon|} \right) \qquad \to \qquad |\varepsilon| = \varepsilon_F \exp \left(-\frac{|\varepsilon_0|}{2\nu V^2} \right). \tag{120}$$

This is essentially the same as the expression (102) for the Kondo temperature. It suggests the following interpretation: the scattering between the conduction electrons and the impurity

spin — the J-term in (100) — becomes significant at temperatures $T \sim T_K$, at which point perturbation theory in J breaks down. This represents the crossover to a state where the impurity spin forms a bound state with the conduction sea, the binding energy of which is $k_B T_K$. This non-magnetic bound state is called the **Kondo singlet**. Put the other way round, one could say that at $T = 0$ the Varma-Yafet solution shows a bound state of binding energy $k_B T_K$, and thus to break that bound state apart and recover the magnetic impurity must require a temperature $T \sim T_K$.

3.6 Recovery of the Fermi liquid at low temperatures

The fact that the susceptibility ceases to diverge below the temperature T_K indicates that the local spin is no longer functioning as a magnetic impurity: it has been **screened** by the conduction sea. Hence the approximate form of the magnetic susceptibility is

$$\chi_{\rm imp}(T) \sim \begin{cases} \dfrac{1}{T} & T > T_K, \\ \\ \text{const.} & T < T_K. \end{cases} \tag{121}$$

So an electron in the conduction sea at temperatures $T \ll T_K$ does not see a magnetic impurity, but rather a non-magnetic point scatterer. These, as we know, do not destabilise the Fermi liquid, and hence the Fermi liquid behaviour is recovered at low temperatures: the magnetic susceptibility once again becomes a constant, and the resistivity recovers its T^2 behaviour. However, the presence of the screened Kondo impurities continues to make itself felt via the Landau parameters; in particular, the effective mass m^* may be much higher than it was at $T \gg T_K$. This is believed to be the key to understanding the so-called 'heavy fermion' materials; we shall discuss these briefly in lecture six.

3.7 Exercises

1. *Magnetic impurities from the Anderson model.* Consider the Anderson Hamiltonian (97) in the case $V = 0$, i.e. where there is no tunnelling between the conduction sea and the impurity. Given that the whole system contains N electrons, write the lowest-energy state in each of the following sectors in second-quantised form:

   (a) impurity empty; all electrons in conduction sea.
   (b) impurity singly occupied; all other electrons in conduction sea.
   (c) impurity doubly occupied; all other electrons in conduction sea.

   Calculate the energies of these states, taking the energy of a filled Fermi sea of N electrons to be E_0, and hence identify the values of the parameters ε_0 and U for which a local moment forms on the impurity level.

2. *Relationship between the Anderson and Kondo models.*

   (a) Identify, in terms of the Anderson model parameters ε_0 and U, the limit in which the singly occupied states have energies infinitely far below the empty and doubly-occupied ones.
   (b) As this limit is approached, one can treat the tunnelling term in (97) perturbatively. Using second-order perturbation theory in V, and comparing your result to the processes described by (100), obtain an expression for J in terms of the parameters ε_0 and U.

3. *The magnetic susceptibility of the Kondo model at low temperatures.* Consider the case of an Anderson impurity in a metallic host, with the addition of an applied magnetic field. The Hamiltonian for this problem is[36]

$$H = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} + \varepsilon_0 \sum_\sigma d^\dagger_\sigma d_\sigma + V \sum_{\mathbf{k}\sigma} \left(c^\dagger_{\mathbf{k}\sigma} d_\sigma + d^\dagger_\sigma c_{\mathbf{k}\sigma} \right) + U d^\dagger_\uparrow d_\uparrow d^\dagger_\downarrow d_\downarrow - \frac{B}{2} \left(d^\dagger_\uparrow d_\uparrow - d^\dagger_\downarrow d_\downarrow \right). \tag{122}$$

   i.e. the Anderson Hamiltonian plus a Zeeman term.

   (a) Construct a Varma-Yafet variational ground state for this problem, and evaluate its binding energy in powers of B up to and including order B^2. (*Hint:* As SU(2) symmetry has now been broken by the magnetic field, you can no longer take the coefficients of the spin-up and spin-down terms in (105) to be equal.)

   (b) Hence obtain the linear magnetic susceptibility of the model (122) at $T = 0$, and find its Kondo limit. Comment on your result.

4 The Hubbard model and the Mott insulator

4.1 In two sentences...

Some materials that the band theory would predict to be metals in fact turn out to be insulators; this insulating state arises from strong correlations, and the resulting state is known as the Mott insulating state. In this lecture we shall study this behaviour by considering building the solid atom by atom, thus deriving a model known as the Hubbard model, and discussing its relation to the Fermi liquid considerations given in previous lectures.

4.2 The tight-binding model

This lecture is about the ways in which a solid can become insulating. One way has already been covered: the Fermi liquid can develop an instability, resulting in a phase transition to a different state which may be an insulator. A second will be dealt with in this section (the 'band insulator'), but applies only when the number of electrons per unit cell is an even integer. A third — arguably the most interesting from the strongly correlated point of view — is the Mott insulator, which will be described in the remainder of this lecture.

To start with, let's take a rather artificial case when we know that a solid ought to be an insulator: namely, when its constituent atoms are extremely far apart. Furthermore, let us for simplicity consider the motion of electrons on such a lattice of atoms in one dimension. Let the operator $c^\dagger_{j\sigma}$ create an electron of spin σ on site j of this lattice, and let the energy of such an electron be ε_0. The state $c^\dagger_{j\sigma} |0\rangle$ (where $|0\rangle$ is the vacuum state) is almost an eigenstate of the problem, but not quite, since there is an amplitude for quantum tunnelling from each site to either of its neighbours[37]. This process is called 'hopping', and its amplitude is called the

[36] Note that, for simplicity, we have applied the magnetic field only to the impurity. This is physically unrealistic, but technically useful; one can show that including a Zeeman term on the conduction electrons does not materially affect the answer.

[37] In fact, there is an amplitude for tunnelling from each site to *any* other. But since the tunnelling amplitude decays exponentially with distance, we make the approximation that nearest-neighbour tunnelling is the only one we need to consider.

hopping integral[38], which we shall denote by $-t$. (The minus sign is chosen more or less conventionally.) The Hamiltonian is therefore

$$H = \sum_{j=1}^{N} \sum_{\sigma} \left(\varepsilon_0 c_{j\sigma}^\dagger c_{j\sigma} - t c_{j\sigma}^\dagger c_{j+1,\sigma} - t c_{j+1,\sigma}^\dagger c_{j\sigma} \right). \tag{123}$$

This is called the **tight-binding model**, since it starts from a basis where each electron is tightly bound to a single lattice site. We have written it for a lattice of N sites, and periodic boundary conditions are assumed, i.e. $c_{N+1,\sigma} \equiv c_{1\sigma}$.

Since this Hamiltonian is translationally invariant, it may be easily diagonalised by Fourier transformation. We introduce the momentum-space c-operators via

$$c_{j\sigma} = \frac{1}{\sqrt{N}} \sum_k e^{ikj} c_{k\sigma}, \tag{124}$$

where the sum over k runs over all the wavenumbers in the first Brillouin zone that are consistent with the boundary conditions. In this case, the lattice spacing is 1, so the first Brillouin zone is defined by[39]

$$-\pi \leqslant k < \pi; \tag{125}$$

and the periodic boundary conditions imply that

$$k = \frac{2\pi n}{N}, \tag{126}$$

where n is an integer. Hence the allowable values of n are[40]

$$n = -\frac{N}{2}, -\frac{N}{2}+1, \ldots, \frac{N}{2}-2, \frac{N}{2}-1. \tag{127}$$

Thus there are N momentum eigenstates, just as there are N position eigenstates, as one would expect.

Substituting (124) into (123), we obtain

$$H = \frac{1}{N} \sum_{kq} \sum_{j=1}^{N} \sum_{\sigma} \left[e^{ij(q-k)} \left(\varepsilon_0 - t e^{iq} - t e^{-ik} \right) c_{k\sigma}^\dagger c_{q\sigma} \right]. \tag{128}$$

We may use the identity

$$\sum_{j=1}^{N} e^{ijx} = N \sum_{n=-\infty}^{\infty} \delta_{x,2\pi n}, \tag{129}$$

to rewrite (128) as[41]

$$H = \sum_{k\sigma} \underbrace{\left(\varepsilon_0 - 2t \cos k \right)}_{=\varepsilon_k} c_{k\sigma}^\dagger c_{k\sigma}. \tag{130}$$

[38]The reason for the term 'integral' is that if you had to calculate t from first-quantised quantum mechanics, you'd do an integral involving the initial and final wavefunctions and the perturbing potential of the lattice of nuclei.

[39]Note carefully that only one of the end-points π and $-\pi$ can be included in the zone, since they correspond to the same physical state.

[40]To avoid definitional problems, let us assume henceforth that N is even.

[41]The eagle-eyed will notice that we have in fact used only the $n = 0$ term on the right-hand side of (129). This is correct, however, since as q and k are restricted to the first Brillouin zone, $q - k$ can never reach $2\pi n$ unless $n = 0$. Thus the $n \neq 0$ terms do not contribute.

This Hamiltonian is now in diagonal form, and the term sitting in front of the number operator is clearly the energy of mode k. Note that eigenstates exist only between energies $\varepsilon_0 - 2t$ (when $k = 0$) and $\varepsilon_0 + 2t$ (when $k = -\pi$). This range of energies is called the **band**, and the difference between the extremal energies ($= 4t$) is called the **bandwidth**. Outside this range of energy there are no modes.

To determine the ground state for N_e electrons, we populate the $2N$ states[42] starting from those of lowest energy, creating a Fermi sea. Evidently if $N_e = 2N$, all states will be filled. In this case there is no Fermi surface, in the sense that no low-energy excitations (indeed, no excitations at all!) are possible: hence the system at $N_e = 2N$ is an insulator. We call it a **band insulator**, since it insulates by virtue of the band's being full.

A significant point about this solution of the tight-binding model is that the single-particle eigenstates are always **delocalised**, i.e. the probability of finding the electron on any lattice site is the same as on any other. To see this, consider the eigenfunction corresponding to a given k-mode:

$$|k\sigma\rangle = c^\dagger_{k\sigma}|0\rangle = \frac{1}{\sqrt{N}}\sum_{j=1}^{N} e^{-ikj}|j\sigma\rangle\,. \tag{131}$$

Thus the probability of finding the electron on site j is given by

$$P_j = \left|\frac{1}{\sqrt{N}}e^{-ikj}\right|^2 = \frac{1}{N}, \tag{132}$$

independent of j. This conclusion holds for arbitrarily small hopping integral t: the band gets narrower in energy, but the eigenstates are always spread over the whole lattice. Therefore, for any number of electrons other than $N_e = 2N$ or the trivial case $N_e = 0$, the model (130) predicts metallic behaviour.

4.3 The Hubbard model and the Mott insulator

There are two problems with the foregoing analysis. Firstly, we experimentally observe many compounds where $N_e \neq 2N$ but which exhibit insulating behaviour: La_2CuO_4, one of the parent compounds of the cuprate superconductors, is a good example. Secondly, the fact that the eigenstates of (130) are all delocalised contradicts our reasoning that the solid should become an insulator when its constituent atoms are sufficiently widely spaced. So what have we got wrong?

There are in fact two possible causes of insulating behaviour that we have omitted. One is disorder: deviations from perfect crystallinity that act as point-like (but non-magnetic) scatterers for the electrons. The presence of disorder destroys translational invariance, invalidating the tight-binding model studied above, and causes some (or in some cases all) of the eigenfunctions to become localised. The fascinating topic of disorder is, however, beyond the scope of this course, and we shall say no more about it[43].

The other neglected feature, as in the previous lecture, is the Coulomb repulsion between electrons when they occupy the same site. This repulsion eventually causes the electrons to 'lock' in place when t gets too small, so we must include it in the Hamiltonian to capture this

[42] The factor of 2 is due, of course, to the fact that each tight-binding eigenstate can be populated by two electrons, one of each spin projection.

[43] The interested reader is referred to T. Brandes and S. Kettemann (eds.), *Anderson Localization and Its Ramifications: Disorder, Phase Coherence, and Electron Correlations* (Springer, 2003) for a lucid and comprehensive treatment.

effect. We do so in the same way that we did for the impurity in the Anderson model: via an additional energy cost U (often called the 'Hubbard U') for doubly occupying a lattice site. The Hamiltonian then becomes[44]

$$H = -t \sum_{j\sigma} \left(c^\dagger_{j\sigma} c_{j+1,\sigma} + c^\dagger_{j+1,\sigma} c_{j\sigma} \right) + U \sum_j c^\dagger_{j\uparrow} c_{j\uparrow} c^\dagger_{j\downarrow} c_{j\downarrow}. \tag{133}$$

This is called the **Hubbard model**. It may easily be generalised to higher dimensions for any lattice geometry:

$$H = -t \sum_{\langle ij \rangle \sigma} \left(c^\dagger_{i\sigma} c_{j\sigma} + c^\dagger_{j\sigma} c_{i\sigma} \right) + U \sum_j c^\dagger_{j\uparrow} c_{j\uparrow} c^\dagger_{j\downarrow} c_{j\downarrow}, \tag{134}$$

where the notation $\langle ij \rangle$ means that the sum runs over all nearest-neighbour pairs of sites on the lattice. As in the Anderson model case, the addition of the U-term transforms the easily solvable tight-binding model into a very hard correlated many-body problem.

Remarkably, an exact solution exists for the one-dimensional Hubbard model (133), but we shall not have time to go into it here; instead, we shall make do with a few elementary observations valid in any number of dimensions. As well as on the parameters U and t (or rather on the ratio U/t), the state of the Hubbard system depends on the filling fraction, defined by

$$f \equiv \frac{N_e}{2N}, \tag{135}$$

where N_e is the number of electrons and N is the number of sites:

- For $f \ll 1$, there are very few electrons in the system, and hence they will not encounter each other very often. Therefore, the U term is largely inoperative, and we expect the state to be metallic.

- For $f = \frac{1}{2}$ — the **half-filling** case — there is one electron per site on average. When the repulsion U is sufficiently large, we therefore expect that the electrons will choose to minimise the effect of the U-term by adopting a state with no double occupancy, i.e. with *exactly* one electron per site[45]. This state is clearly an insulator, since to move charge one needs to change the occupancy of a site, which always costs energy U. It is called the **Mott insulating state**. An elementary argument (certainly not a proof!) that the state does become insulating for $f = \frac{1}{2}$ is given in the Exercises.

- For $1 - f \ll 1$, where the system is nearly full of electrons, we may think of it as a few holes inserted into a filled band. Introducing the representation

$$h^\dagger_{j\sigma} \equiv c_{j\sigma}, \tag{136}$$

so that the operator $h^\dagger_{j\sigma}$ creates a *hole* at site j with spin σ, the Hamiltonian (133) becomes

$$H = -t \sum_{\langle ij \rangle \sigma} \left(h_{i\sigma} h^\dagger_{j\sigma} + h_{j\sigma} h^\dagger_{i\sigma} \right) + U \sum_j h_{j\uparrow} h^\dagger_{j\uparrow} h_{j\downarrow} h^\dagger_{j\downarrow}. \tag{137}$$

[44] Without loss of generality, we have chosen to set $\varepsilon_0 = 0$.

[45] In fact, the Hubbard model on a hypercubic lattice is always insulating at $f = \frac{1}{2}$ for any non-zero U, in any number of dimensions. This is because the associated tight-binding model has a perfectly nested Fermi surface, and hence a magnetic instability kicks in first. This is one of the problems with the Mott insulating transition: it's very hard to observe because an itinerant-type magnetic instability often pre-empts it.

Noting that the h-operators obey the same anticommutation rules as the c-operators, we may manipulate this Hamiltonian as follows:

$$\begin{aligned} H &= t\sum_{\langle ij\rangle\sigma}\left(h^\dagger_{j\sigma}h_{i\sigma}+h^\dagger_{i\sigma}h_{j\sigma}\right)+U\sum_j\left(1-h^\dagger_{j\uparrow}h_{j\uparrow}\right)\left(1-h^\dagger_{j\downarrow}h_{j\downarrow}\right) \\ &= t\sum_{\langle ij\rangle\sigma}\left(h^\dagger_{j\sigma}h_{i\sigma}+h^\dagger_{i\sigma}h_{j\sigma}\right)-U\sum_{j\sigma}h^\dagger_{j\sigma}h_{j\sigma}+U\sum_j h^\dagger_{j\uparrow}h_{j\uparrow}h^\dagger_{j\downarrow}h_{j\downarrow}, \end{aligned} \tag{138}$$

where in the last line we have dropped an evidently constant term. Furthermore, for fixed f the second term is also constant, leaving only the first and third terms physically relevant. Hence the holes also obey a Hubbard model (with the sign of the hopping t reversed). Therefore, if we define the hole filling factor f_h by

$$f_h \equiv 1-f, \tag{139}$$

we expect the system at filling f_h to have the same physics as the system at f, but with the charge carried by holes instead of by electrons. Therefore we expect a metallic state also at fillings close to 1.

Hence, as the filling is increased from 0 to 1, we generically expect two transitions: a transition from the low-f metallic state to the Mott insulator near $f=\frac{1}{2}$, and another from the Mott insulator back to the (hole-like) metallic state at high fillings. In one dimension, we know from the exact solution that both transitions happen *exactly at* half-filling, i.e. the model is metallic except when there is *precisely* one electron per site. What happens in higher dimensions is still an open question.

4.4 Magnetism in the Mott state

The attentive reader will have noticed that our description of the Mott state is not unique. By stipulating only that there be exactly one electron per site, we have failed to specify the spin states adopted by those electrons. We now turn to that question. Before attempting a rigorous derivation, let us discuss physically how this must turn out. We are essentially trying to decide whether neighbouring spins should be parallel (ferromagnetism) or antiparallel (antiferromagnetism). If they are parallel, no hopping can take place, since the exclusion principle forbids two parallel-spin electrons from existing on the same site. If they are anti-parallel, however, virtual hopping fluctuations can take place. Since virtual fluctuations always lower the energy, the anti-parallel state is preferred, and hence we expect the resulting state to be **antiferromagnetic**.

Let's prove that. Since the Mott state is an eigenstate of the U-term in (134), with eigenvalue zero independently of the spin configuration, it must be the t-term that is responsible for determining the spin ordering that is energetically preferred. However, a single action of the t-term on the Mott state will result in a state where one site is doubly occupied, with energy cost U. Two consecutive operations of the t-term, on the other hand, would permit a return to a state with exactly one electron per site (though not necessarily with the same spin configuration). Hence we are led to consider second-order perturbation theory in t — or, more strictly, in t/U, which we take as a small parameter.

Let us call the two terms in the Hubbard model (134) H_t and H_U respectively. Then the

effective Hamiltonian in the case $U \gg t$ is given by[46]

$$H_{\text{eff}} = PH_t(1-P)\frac{1}{E_0 - H_U}(1-P)H_tP \tag{140}$$

to order t^2. The operator P projects into the (highly degenerate) manifold of ground states of H_U, and E_0 is the energy of that ground-state manifold (which in this model is zero).

Let us consider the possible intermediate states attainable by an action of H_t on a ground state spin configuration. If i and j are neighbouring sites, then H_t can remove an electron from site i and put it on site j, provided that the spin orientations are antiparallel. So the first action of H_t is to effect the transition

$$|\sigma\rangle_i \otimes |\bar{\sigma}\rangle_j \to |\sigma\bar{\sigma}\rangle_i \otimes |0\rangle_j \,. \tag{141}$$

where $\bar{\sigma}$ denotes the opposite spin to spin σ. This intermediate state has energy U. The second action of H_t must return one of those electrons to site j, but it may be either of them; therefore the transition effected is

$$|\sigma\bar{\sigma}\rangle_i \otimes |0\rangle_j \to |\sigma\rangle_i \otimes |\bar{\sigma}\rangle_j + |\bar{\sigma}\rangle_i \otimes |\sigma_j\rangle \,. \tag{142}$$

So the overall action of H_{eff}, including the t coefficients and the intermediate-state energy cost, is

$$H_{\text{eff}}\left(|\sigma\rangle_i \otimes |\bar{\sigma}\rangle_j\right) = -\frac{t^2}{U}\left(|\sigma\rangle_i \otimes |\bar{\sigma}\rangle_j + |\bar{\sigma}\rangle_i \otimes |\sigma_j\rangle\right) \tag{143}$$

$$H_{\text{eff}}\left(|\sigma\rangle_i \otimes |\sigma\rangle_j\right) = 0. \tag{144}$$

It is not hard to show that the following Hamiltonian has exactly these matrix elements:

$$H = J\sum_{\langle ij\rangle} \mathbf{S}_i \cdot \mathbf{S}_j, \tag{145}$$

where $\mathbf{S}_i$ is a vector of spin-$\frac{1}{2}$ operators on site i, and $J \equiv \frac{8t^2}{U}$. This is called the **Heisenberg model**, and it describes the magnetic interactions deep in the Mott insulating state.

4.5 Exercises

1. *The two-site Hubbard model.* This problem concerns the two-site Hubbard model, which you will solve exactly. Its Hamiltonian is

$$H = -2t\sum_{\sigma}\left(c^\dagger_{1\sigma}c_{2\sigma} + c^\dagger_{2\sigma}c_{1\sigma}\right) + U\sum_{j=1}^{2} c^\dagger_{j\uparrow}c_{j\uparrow}c^\dagger_{j\downarrow}c_{j\downarrow}. \tag{146}$$

   (a) Why is the coefficient of the hopping term $-2t$ rather than $-t$?

   (b) Construct the full basis of states for the model (146) using the position-space creation operators,

$$|0\rangle \,, c^\dagger_{1\uparrow}|0\rangle \,, c^\dagger_{1\uparrow}c^\dagger_{2\downarrow}|0\rangle \,, \text{etc.} \tag{147}$$

   How many are there? (*Hint:* Make sure that all states in the set you write down are linearly independent.)

[46] For a derivation of this useful operator-based version of perturbation theory, see, for example, I. Lindgren, *J. Phys. B: Atom. Molec. Phys.* **7**, 2441 (1974).

(c) Show that the total number of electrons

$$N_e \equiv \sum_{j\sigma} c^\dagger_{j\sigma} c_{j\sigma} \tag{148}$$

and the total z-projection of the spin

$$S^z_T \equiv \frac{1}{2} \sum_j \left(c^\dagger_{j\uparrow} c_{j\uparrow} - c^\dagger_{j\downarrow} c_{j\downarrow} \right) \tag{149}$$

are both conserved quantities of the Hamiltonian (146). Hence divide the space of states spanned by the basis (147) into subspaces of fixed N_e and S^z_T.

(d) Diagonalise H within each of these subspaces, and hence obtain all energy eigenvalues and eigenstates of the model (146).

(e) Analyse your solution in the particular cases: (i) $U \ll t$; (ii) $U \gg t$. Pay particular attention to the $N_e = 2$ subspace.

2. *The Heisenberg model.* Prove that the matrix elements of (140) and (145) are the same provided that they operate within the ground state manifold of H_U.

5 Magnetism in Mott insulators: the Heisenberg model analysed

5.1 In two sentences...

Having established that nearest-neighbour spins in the Mott state have a tendency to align antiferromagnetically, there remains the question of the global magnetic state to which this leads. One option is the Néel state (…-up-down-up-down-up-down-…), but thermal or quantum fluctuations can destroy this, in the latter case perhaps leading to more interesting forms of order; whether this happens turns out to depend on the dimensionality of the lattice.

5.2 Classical magnetism

To determine the magnetic ground state of the Heisenberg model, it is convenient to begin by obtaining the ground state magnetic ordering pattern[47] of a classical magnet on the same lattice.

By "magnet" we mean a system of spins on the sites of a lattice. The spins are represented by spin operators; the spin operator on site j will be denoted

$$\hat{\mathbf{S}}_j = \begin{pmatrix} \hat{S}^x_j \\ \hat{S}^y_j \\ \hat{S}^z_j \end{pmatrix}. \tag{150}$$

The commutation relation between spin operators at sites j and k is given by

$$\left[\hat{S}^a_j, \hat{S}^b_k \right] = i\hbar \epsilon^{abc} \delta_{jk} \hat{S}^c_j, \tag{151}$$

[47] Or patterns—there may be many degenerate ground state orderings.

where ϵ^{abc} is the Levi-Civita symbol, which is defined by

$$\epsilon^{abc} \equiv \begin{cases} 1 & abc \text{ is a cyclic permutation of } xyz \\ -1 & abc \text{ is an anti-cyclic permutation of } xyz \\ 0 & \text{otherwise.} \end{cases} \tag{152}$$

Note that spin operators on different sites commute with each other. The magnitude of the spin on each site is S, which means that[48]

$$\hat{\mathbf{S}}_j \cdot \hat{\mathbf{S}}_j = S(S+1)\hbar^2. \tag{153}$$

By "classical magnet", we mean a magnet considered in the $S \to \infty$ limit. To see what happens in this limit, let us rewrite the spin operator as a magnitude times an orientation:

$$\hat{\mathbf{S}}_j \equiv \hbar S \, \hat{\mathbf{s}}_j. \tag{154}$$

Substituting this definition into the commutation relation (151), we find that

$$\left[\hat{s}_j^a, \hat{s}_k^b\right] = \frac{i}{S}\delta_{jk}\epsilon^{abc}\hat{s}_j^c. \tag{155}$$

In the limit $S \to \infty$, the right hand side of this commutation relation becomes zero. The spin operators therefore lose their quantum properties (which are all to do with the operators' failure to commute), and become simply a set of unit vectors $\mathbf{s}_j$ on the sites j of the lattice[49]; since they are unit vectors, they satisfy

$$\mathbf{s}_j \cdot \mathbf{s}_j = 1. \tag{156}$$

The Heisenberg Hamiltonian, derived from the $U \gg t$ Hubbard model in the previous lecture, is given by

$$H = J \sum_{\langle jk \rangle} \hat{\mathbf{S}}_j \cdot \hat{\mathbf{S}}_k. \tag{157}$$

If $J < 0$, the Hamiltonian favours parallel spins, and the interaction is called *ferromagnetic* (or FM)[50]. If $J > 0$, it favours antiparallel spins, and the interaction is called *antiferromagnetic* (or AFM).

In the $S \to \infty$ limit, the Hamiltonian (157) becomes simply

$$H = \hbar^2 J S^2 \sum_{\langle jk \rangle} \mathbf{s}_j \cdot \mathbf{s}_k. \tag{158}$$

If we define a rescaled interaction constant $J_{\mathrm{cl}} \equiv \hbar^2 J S^2$, the Hamiltonian takes on the simple form

$$H = J_{\mathrm{cl}} \sum_{\langle jk \rangle} \mathbf{s}_j \cdot \mathbf{s}_k. \tag{159}$$

[48] More precisely, it means that all of the states that we shall consider are eigenstates of the operator $\hat{\mathbf{S}}_j \cdot \hat{\mathbf{S}}_j$, with eigenvalue $S(S+1)\hbar^2$.

[49] Since this is no longer a quantum problem, we have dropped the "hat" from these vectors. Another way of viewing this is as follows: Since all operators now commute with all others, they can be simultaneously diagonalised. This means that we can work in a basis set where the states are eigenstates of all the operators $\hat{\mathbf{s}}_j$. But if this is true, then we can simply replace the operators by their eigenvalues in all our equations—i.e. we can remove the "hat".

[50] The nearest-neighbour Hubbard model, as we saw above, always gives a positive J. But models for more complicated lattices composed of more than one element may yield negative J, so we consider both cases.

It is convenient to "unpack" the nature of the sum in a more mathematically explicit way. To do this, notice that the sum includes each "bond" on the lattice (i.e. each line that joins a pair of nearest neighbours) precisely once. Another way of obtaining each bond is to start at a given lattice site j, and to link it to its nearest neighbours by a set of nearest-neighbour vectors[51] $\mathbf{m}$. This counts all bonds twice, so a correcting factor of $\frac{1}{2}$ is needed. The Hamiltonian may thus be written

$$H = \frac{J_{\rm cl}}{2} \sum_j \sum_{\mathbf{m}} \mathbf{s}_j \cdot \mathbf{s}_{j+\mathbf{m}}, \tag{160}$$

where the site label $j + \mathbf{m}$ refers to the site at which one arrives by starting at site j (whose position is $\mathbf{x}_j$) and translating by the vector $\mathbf{m}$.

The remainder of this section shows how to obtain the ground state ordering pattern of the classical magnet described by this Hamiltonian. Subsection 5.2.1 shows how to rewrite the Hamiltonian in a more useful form, and subsection 5.2.2 shows how to find the ground state ordering pattern, and illustrates this using two simple examples.

5.2.1 Fourier (or spin-wave) representation

To see how to simplify the Hamiltonian (160), we should think about the symmetries of the system. As with the tight-binding model above, the presence of lattice translational invariance prompts us to transform the operators into momentum space:

$$\mathbf{s}_j \equiv \frac{1}{\sqrt{N}} \sum_{\mathbf{p}} e^{i\mathbf{p}\cdot\mathbf{x}_j} \mathbf{s}_{\mathbf{p}}. \tag{161}$$

In this formula, N is the number of sites of the lattice, and $\mathbf{p}$ runs over all values consistent with the boundary conditions and lying in the first Brillouin zone. The vector $\mathbf{s}_{\mathbf{p}}$ represents the Fourier component of the magnetisation with wavevector $\mathbf{p}$: a **spin wave**. Substituting (161) into the classical Hamiltonian (160), we obtain

$$\begin{aligned} H &= \frac{J_{\rm cl}}{2} \sum_j \sum_{\mathbf{m}} \mathbf{s}_j \cdot \mathbf{s}_{j+\mathbf{m}} \\ &= \frac{J_{\rm cl}}{2N} \sum_j \sum_{\mathbf{m}} \sum_{\mathbf{p},\mathbf{q}} e^{i\mathbf{p}\cdot\mathbf{x}_j} e^{i\mathbf{q}\cdot(\mathbf{x}_j+\mathbf{m})} \mathbf{s}_{\mathbf{p}} \cdot \mathbf{s}_{\mathbf{q}} \\ &= \sum_{\mathbf{p},\mathbf{q}} \left\{ \frac{1}{N} \sum_j e^{i\mathbf{x}_j\cdot(\mathbf{p}+\mathbf{q})} \right\} \left\{ \frac{J_{\rm cl}}{2} \sum_{\mathbf{m}} e^{i\mathbf{q}\cdot\mathbf{m}} \right\} \mathbf{s}_{\mathbf{p}} \cdot \mathbf{s}_{\mathbf{q}}. \end{aligned} \tag{162}$$

At first glance, this does not appear to have made things much better: we've replaced two sums by four! However, the Hamiltonian simplifies considerably if we use the identity

$$\frac{1}{N} \sum_j e^{i\mathbf{p}_0\cdot\mathbf{x}_j} = \sum_{\mathbf{G}} \delta_{\mathbf{p}_0,\mathbf{G}}, \tag{163}$$

where the sum runs over all vectors $\mathbf{G}$ of the reciprocal lattice. This statement, which is the multi-dimensional generalisation of (129), is sometimes known as **Bloch's Theorem**. Substituting (163) into the Hamiltonian (162), and noting as before that only the contribution of the

[51] For example, on the two-dimensional square lattice the nearest-neighbour vectors would be given by $\mathbf{m} = \begin{pmatrix} a \\ 0 \end{pmatrix}, \begin{pmatrix} -a \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ a \end{pmatrix}, \begin{pmatrix} 0 \\ -a \end{pmatrix}$, where a is the lattice spacing.

$\mathbf{G} = \mathbf{0}$ term needs to be retained, we obtain

$$\begin{aligned} H &= \sum_{\mathbf{p},\mathbf{q}} \delta_{\mathbf{p}+\mathbf{q},\mathbf{0}} \left\{ \frac{J_{\rm cl}}{2} \sum_{\mathbf{m}} e^{i\mathbf{q}\cdot\mathbf{m}} \right\} \mathbf{s}_{\mathbf{p}} \cdot \mathbf{s}_{\mathbf{q}} \\ &= \sum_{\mathbf{p}} \gamma_{\mathbf{p}} \mathbf{s}_{\mathbf{p}} \cdot \mathbf{s}_{-\mathbf{p}}, \end{aligned} \tag{164}$$

where we have defined

$$\gamma_{\mathbf{p}} \equiv \frac{J_{\rm cl}}{2} \sum_{\mathbf{m}} e^{-i\mathbf{p}\cdot\mathbf{m}}. \tag{165}$$

This all-important quantity is known as the *structure factor*[52].

To think about the ground states of (164), it is important to note that the various $\mathbf{s}_{\rm p}$ are not independent of each other. The reason is that the original spins are constrained to be of unit length:

$$\mathbf{s}_j \cdot \mathbf{s}_j = 1 \qquad \text{for all } j. \tag{166}$$

Inserting the Fourier decomposition of $\mathbf{s}_j$, we find that

$$\frac{1}{N} \sum_{\mathbf{k},\mathbf{p}} e^{i(\mathbf{k}+\mathbf{p})\cdot\mathbf{x}_j} \mathbf{s}_{\mathbf{k}} \cdot \mathbf{s}_{\mathbf{p}} = 1 \qquad \text{for all } j. \tag{167}$$

This set of constraints is in general quite difficult to satisfy, although for the simple examples we shall consider it's not too bad. For now, we need only note that (166) implies that the vectors $\mathbf{s}_j$ are real (obviously!), and it can easily be shown that

$$\mathbf{s}_j = \mathbf{s}_j^* \qquad \longrightarrow \qquad \mathbf{s}_{-\mathbf{p}} = \mathbf{s}_{\mathbf{p}}^* \tag{168}$$

(where the star denotes complex conjugation). Substituting this into the Hamiltonian (164) we obtain

$$H = \sum_{\mathbf{p}} \gamma_{\mathbf{p}} \left| \mathbf{s}_{\mathbf{p}} \right|^2 . \tag{169}$$

Since the positive and negative values of $\mathbf{p}$ are not independent, it makes sense to rewrite this Hamiltonian as a sum over only half of $\mathbf{p}$-space:[53]

$$H = \sum_{\mathbf{p} \text{ s.t. } p_x \geqslant 0} \left[\frac{1}{2} \left(\gamma_{\mathbf{p}} + \gamma_{-\mathbf{p}} \right) \right] \left| \mathbf{s}_{\mathbf{p}} \right|^2 . \tag{170}$$

5.2.2 Obtaining the classical ground state(s)

To find the classical ground state, we simply need to find the spin configuration that minimises the total energy, i.e. the Hamiltonian (170), subject to the constraints (167). As mentioned above, this is in principle quite a tricky business. However, the intuitive solution of just lumping all the weight into the $\mathbf{p}$-mode with the lowest energy is often sufficient to determine the classical order. Specifically, one can try the solution

$$\mathbf{s}_{\mathbf{p}} = \begin{cases} \sqrt{N}\, \mathbf{n}_0 & \mathbf{p} = \mathbf{p}_0, \\ 0 & \text{otherwise}, \end{cases} \tag{171}$$

[52] It is sometimes defined without the leading factors of $J_{\rm cl}$, $\frac{1}{2}$, or both, and indeed sometimes with other constants in front of it. But clearly its functional dependence on $\mathbf{p}$ is the important thing.

[53] Which half, of course, is arbitrary. We have chosen to split $\mathbf{p}$-space in two by cutting it along the p_y-axis, but this choice is purely conventional. Note that in many cases $\gamma_{\mathbf{p}} = \gamma_{-\mathbf{p}}$, whereupon (170) simplifies even further.

where $\mathbf{p}_0$ is defined as the wavevector that minimises $\frac{1}{2}(\gamma_{\mathbf{p}} + \gamma_{-\mathbf{p}})$, and $\mathbf{n}_0$ is a unit vector describing the global orientation of the magnetic order. (Since the Hamiltonian only specifies the *relative* orientations of the spins, there will always be one spin whose direction we are completely free to choose. This spin points in the direction $\mathbf{n}_0$.) Then this expression for $\mathbf{s}_{\mathbf{p}}$ may be substituted back into (161) to obtain the ground state magnetisation pattern.

To close the section on classical magnetism, let us give a couple of simple examples:

- Heisenberg ferromagnet on the 2D square lattice.

  In this case, there are four nearest-neighbour vectors $\mathbf{m}$, given by

  $$\mathbf{m} = \begin{pmatrix} a \\ 0 \end{pmatrix}, \quad \begin{pmatrix} -a \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ a \end{pmatrix}, \quad \begin{pmatrix} 0 \\ -a \end{pmatrix}, \tag{172}$$

  where a is the lattice spacing. A simple calculation shows that

  $$\gamma_{\mathbf{p}} = \gamma_{-\mathbf{p}} = J_{\mathrm{cl}} \left[\cos(p_x a) + \cos(p_y a) \right]. \tag{173}$$

  Differentiation with respect to p_x and p_y shows that this function has four extrema, but only one minimum. In the case where J_{cl} is negative (the ferromagnetic case), the minimum is given by

  $$\mathbf{p}_0 \equiv (p_{0x}, p_{0y}) = (0, 0). \tag{174}$$

  Hence, in accordance with the procedure outlined above, we set

  $$\mathbf{s}_{\mathbf{p}} = \begin{cases} \sqrt{N}\, \mathbf{n}_0 & \mathbf{p} = \mathbf{0}, \\ 0 & \text{otherwise.} \end{cases} \tag{175}$$

  Substituting this into (161), we find that

  $$\mathbf{s}_j = \mathbf{n}_0 \tag{176}$$

  independently of j. This simply means that all spins point in the same (arbitrary) direction $\mathbf{n}_0$, which is precisely what one would expect for a classical ferromagnet.

- Heisenberg antiferromagnet on the 2D square lattice.

  The calculation for the antiferromagnetic case gives precisely the same $\gamma_{\mathbf{p}}$ as we obtained above in equation (173). However, now J_{cl} is positive, and so maxima and minima are interchanged compared to the ferromagnetic problem. The new minimum, as may be easily shown, is given by

  $$\mathbf{p}_0 = \left(-\frac{\pi}{a}, -\frac{\pi}{a} \right), \tag{177}$$

  and hence we choose

  $$\mathbf{s}_{\mathbf{p}} = \begin{cases} \sqrt{N}\, \mathbf{n}_0 & \mathbf{p} = \left(-\frac{\pi}{a}, -\frac{\pi}{a} \right), \\ 0 & \text{otherwise.} \end{cases} \tag{178}$$

  Substituting this into the expression (161) for $\mathbf{s}_j$ gives

  $$\begin{aligned} \mathbf{s}_j &= \frac{1}{\sqrt{N}} \sum_{\mathbf{p}} e^{i\mathbf{p}\cdot\mathbf{x}_j} \mathbf{s}_{\mathbf{p}} \\ &= e^{-i\left(\frac{\pi}{a}, \frac{\pi}{a}\right)\cdot\mathbf{x}_j} \mathbf{n}_0. \end{aligned} \tag{179}$$

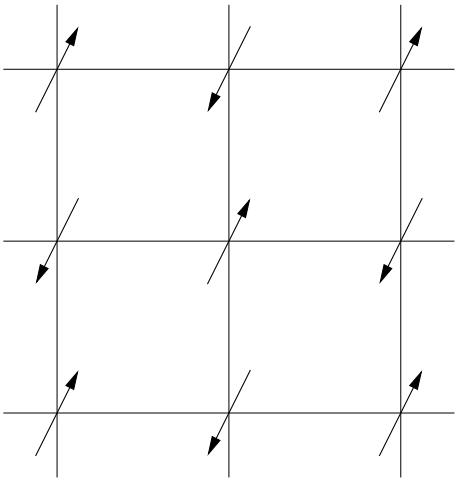


Figure 7: The Néel state for a classical 2D square lattice antiferromagnet. The spins on one sublattice point in the direction $\mathbf{n}_0$; those on the other point in the direction $-\mathbf{n}_0$.

We may write the position vector $\mathbf{x}_j$ as

$$\mathbf{x}_j = \begin{pmatrix} ma \\ na \end{pmatrix} \tag{180}$$

where n and m are integer co-ordinates of the lattice site j, and hence

$$\mathbf{s}_j = e^{-i\pi(n+m)}\mathbf{n}_0 = (-1)^{n+m}\,\mathbf{n}_0. \tag{181}$$

This produces the magnetisation pattern shown in Fig. 7, which is usually called the *Néel state*. This is, of course, precisely what one would expect for an antiferromagnet: each spin is antiparallel to all of its nearest neighbours.

In both of these cases, the classical ground state was fairly easy to guess, which is reflected in the fact that the structure factor exhibited only a single minimum. For more complicated lattices — in particular those featuring polygons with an odd number of sides — things are not so simple. Such lattices are said to be **geometrically frustrated**; perhaps the simplest example, the triangular lattice, appears in the Exercises.

5.3 Semi-classical magnetism (spin wave theory)

Having dealt with the procedure for obtaining the classical ($S \to \infty$) ground states of Heisenberg magnets, and provided a couple of simple examples, we now proceed to reincorporate quantum effects by considering S to be finite (though still large). This suggests the possibility of an expansion in powers of the parameter $1/S$. The key to achieving this lies in finding a suitable way to represent the spin operators $\hat{\mathbf{S}}_j$.

Subsection 5.3.1 shows how to write down an approximate representation for the spin operators $\hat{\mathbf{S}}_j$ in terms of boson creation and annihilation operators $\hat{a}_j^\dagger$ and $\hat{a}_j$. (This representation is due to Holstein and Primakoff.) Subsection 5.3.2 then uses the Holstein-Primakoff transformation to solve the problem of spin waves in a 3D Heisenberg ferromagnet. Finally, in subsection 5.3.3, we see how to check the consistency of the $1/S$ expansion.

5.3.1 The Holstein-Primakoff transformation

Let us consider how to tackle the problem of a quantum Heisenberg magnet in the case where the spin S is large, but not infinite. For simplicity, let us study the nearest-neighbour Heisenberg ferromagnet on a d-dimensional lattice; the Hamiltonian is

$$\begin{aligned}\hat{H} &= -|J| \sum_{\langle jk \rangle} \hat{\mathbf{S}}_j \cdot \hat{\mathbf{S}}_k \\ &= -\frac{|J|}{2} \sum_j \sum_{\mathbf{m}} \left[\hat{S}_j^x \hat{S}_{j+\mathbf{m}}^x + \hat{S}_j^y \hat{S}_{j+\mathbf{m}}^y + \hat{S}_j^z \hat{S}_{j+\mathbf{m}}^z \right], \end{aligned} \tag{182}$$

where we have expanded the dot product and, as above, rewritten the sum using the nearest-neighbour vectors $\mathbf{m}$. The spin raising and lowering operators on site j are given by

$$\hat{S}_j^{\pm} = \hat{S}_j^x \pm i \hat{S}_j^y; \tag{183}$$

in terms of these, the Hamiltonian becomes

$$\hat{H} = -\frac{|J|}{2} \sum_j \sum_{\mathbf{m}} \left[\hat{S}_j^z \hat{S}_{j+\mathbf{m}}^z + \frac{1}{2} \left(\hat{S}_j^+ \hat{S}_{j+\mathbf{m}}^- + \hat{S}_j^- \hat{S}_{j+\mathbf{m}}^+ \right) \right]. \tag{184}$$

Now, we know that in the $S \to \infty$ limit the ground state ordering pattern is one in which all spins point along the same direction. Let us choose that direction to be given by the positive z axis. The essence of spin-wave theory is to assume that the introduction of quantum effects does not qualitatively change the nature of this ordering: we shall see later how to check whether this assumption is correct. Mathematically, we proceed by thinking of the problem as follows: When the spin-projection on a given site is $\hbar S$, we say that it is 'empty'. Every time the spin is lowered, we describe it by the creation of a boson on the site; every time a spin is raised, we describe it by the annihilation of a boson. Thus we construct a mapping between the spin-projections and the number of bosons on a given site:

$S^z/\hbar$	S	$S-1$	$S-2$	$S-3$	...	$-S$
n	0	1	2	3	...	$2S$

In terms of operators, we can write this mapping as

$$\begin{aligned}\hat{S}^z &= \hbar \left(S - \hat{a}^\dagger \hat{a} \right), \\ \hat{S}^- &= \hbar \hat{a}^\dagger \left(2S - \hat{a}^\dagger \hat{a} \right)^{1/2}, \\ \hat{S}^+ &= \hbar \left(2S - \hat{a}^\dagger \hat{a} \right)^{1/2} \hat{a}. \end{aligned} \tag{185}$$

It is called the **Holstein-Primakoff transformation**. The extra factor $\left(2S - \hat{a}^\dagger \hat{a}\right)^{1/2}$ is necessary to reproduce the correct commutation relations between the spin operators; physically, it may be thought of as ensuring that one cannot create more than $2S$ bosons on a given site, i.e. that one cannot lower the spin more than $2S$ times.

The presence of this extra factor is an irritation, since square roots of operators are generally hard to handle. But if we are interested only in large values of S, then we may approximate the factor (very crudely!) by

$$\left(2S - \hat{a}^\dagger \hat{a}\right)^{1/2} \approx \sqrt{2S}, \tag{186}$$

which should be all right provided that the number of bosons on each site remains small compared to $2S$. (We shall check this assumption later.) Those who make the simplification (186) are said to be doing **linearised spin-wave theory**.

Substituting the linearised Holstein-Primakoff expressions into the Hamiltonian (184), we find that

$$\begin{aligned}\hat{H} &= -\frac{|J|}{2}\sum_j\sum_{\mathbf{m}}\left[\hat{S}_j^z\hat{S}_{j+\mathbf{m}}^z+\frac{1}{2}\left(\hat{S}_j^+\hat{S}_{j+\mathbf{m}}^-+\hat{S}_j^-\hat{S}_{j+\mathbf{m}}^+\right)\right]\\ &= -\frac{|J|}{2}\sum_j\sum_{\mathbf{m}}\left[S^2-S\left(\hat{a}_j^\dagger\hat{a}_j+\hat{a}_{j+\mathbf{m}}^\dagger\hat{a}_{j+\mathbf{m}}-\hat{a}_{j+\mathbf{m}}^\dagger\hat{a}_j-\hat{a}_j^\dagger\hat{a}_{j+\mathbf{m}}\right)\right]+O(1)\\ &= -\frac{\hbar^2|J|Nz}{2}S^2+\frac{\hbar^2|J|S}{2}\sum_{j,\mathbf{m}}\left(\hat{a}_j^\dagger\hat{a}_j+\hat{a}_{j+\mathbf{m}}^\dagger\hat{a}_{j+\mathbf{m}}-\hat{a}_{j+\mathbf{m}}^\dagger\hat{a}_j-\hat{a}_j^\dagger\hat{a}_{j+\mathbf{m}}\right)+O(1).\end{aligned} \tag{187}$$

We have performed the sum in the first term, using z to represent the number of nearest neighbours that each spin has: this is called the **co-ordination number** of the lattice. The first term may be recognised as the classical energy of the ferromagnetic state, which we shall henceforth write simply as $E_{\rm cl}$.

5.3.2 Diagonalising the Hamiltonian for a simple ferromagnet

We are now faced with the problem of representing the Hamiltonian

$$\hat{H}=E_{\rm cl}+\frac{\hbar^2|J|S}{2}\sum_{j,\mathbf{m}}\left(\hat{a}_j^\dagger\hat{a}_j+\hat{a}_{j+\mathbf{m}}^\dagger\hat{a}_{j+\mathbf{m}}-\hat{a}_{j+\mathbf{m}}^\dagger\hat{a}_j-\hat{a}_j^\dagger\hat{a}_{j+\mathbf{m}}\right) \tag{188}$$

in a manageable form. As in the classical case, our Hamiltonian is invariant under lattice translations, so the trick for doing this is the same as above: to Fourier-transform the operators $\hat{a}_j$. So we write[54]

$$\hat{a}_j=\frac{1}{\sqrt{N}}\sum_{\mathbf{p}}e^{i\mathbf{p}\cdot\mathbf{x}_j}\hat{a}_{\mathbf{p}}, \tag{189}$$

and by conjugating this relation we obtain that

$$\hat{a}_j^\dagger=\frac{1}{\sqrt{N}}\sum_{\mathbf{p}}e^{-i\mathbf{p}\cdot\mathbf{x}_j}\hat{a}_{\mathbf{p}}^\dagger. \tag{190}$$

Substituting these into (188), we find that

$$\hat{H}=E_{\rm cl}+\frac{\hbar^2|J|S}{2N}\sum_{j,\mathbf{m}}\sum_{\mathbf{p},\mathbf{q}}e^{i\mathbf{x}_j\cdot(\mathbf{q}-\mathbf{p})}\left[1+e^{i\mathbf{m}\cdot(\mathbf{q}-\mathbf{p})}-e^{i\mathbf{m}\cdot\mathbf{q}}-e^{-i\mathbf{m}\cdot\mathbf{p}}\right]\hat{a}_{\mathbf{p}}^\dagger\hat{a}_{\mathbf{q}}. \tag{191}$$

[54]The analysis of the allowable values of $\mathbf{p}$ proceeds similarly to that given in the classical case. In particular, the notion of choosing values of $\mathbf{p}$ to lie in the first Brillouin zone still applies.

As before, we may use the result (163) to perform the sum over j, obtaining

$$\begin{aligned}\hat{H} &= E_{\rm cl} + \frac{\hbar^2|J|S}{2}\sum_{\mathbf{m}}\sum_{\mathbf{p},\mathbf{q}}\delta_{\mathbf{p}-\mathbf{q},\mathbf{0}}\left[1 + e^{i\mathbf{m}\cdot(\mathbf{q}-\mathbf{p})} - e^{i\mathbf{m}\cdot\mathbf{q}} - e^{-i\mathbf{m}\cdot\mathbf{p}}\right]\hat{a}^\dagger_{\mathbf{p}}\hat{a}_{\mathbf{q}} \\ &= E_{\rm cl} + \frac{\hbar^2|J|S}{2}\sum_{\mathbf{m}}\sum_{\mathbf{p}}\left[2 - e^{i\mathbf{m}\cdot\mathbf{p}} - e^{-i\mathbf{m}\cdot\mathbf{p}}\right]\hat{a}^\dagger_{\mathbf{p}}\hat{a}_{\mathbf{p}} \\ &= E_{\rm cl} + \sum_{\mathbf{p}}\varepsilon_{\mathbf{p}}\hat{n}_{\mathbf{p}}, \end{aligned} \tag{192}$$

where $\hat{n}_{\mathbf{p}}$ is the number of bosons in the state with wavevector $\mathbf{p}$, and $\varepsilon_{\mathbf{p}}$ is given by

$$\varepsilon_{\mathbf{p}} = \hbar^2|J|S\sum_{\mathbf{m}}\left[1 - \cos\left(\mathbf{p}\cdot\mathbf{m}\right)\right]. \tag{193}$$

We see from the form of the Hamiltonian that $\varepsilon_{\mathbf{p}}$ is simply the energy of a spin wave[55] with wavevector $\mathbf{p}$. Notice that the form of $\varepsilon_{\mathbf{p}}$ is essentially the same as the structure factor $\gamma_{\mathbf{p}}$. Thus the same function determines both the classical ordering (by where its minima are) and the quantum excitations (by its form near the minimum)[56].

For the specific case of the three-dimensional cubic lattice, the six nearest-neighbour vectors $\mathbf{m}$ are

$$\mathbf{m} = \begin{pmatrix}\pm a\\0\\0\end{pmatrix}, \quad \begin{pmatrix}0\\\pm a\\0\end{pmatrix}, \quad \begin{pmatrix}0\\0\\\pm a\end{pmatrix}, \tag{194}$$

where a is the lattice spacing. Hence $\varepsilon_{\mathbf{p}}$ is given by

$$\varepsilon_{\mathbf{p}} = 2\hbar^2|J|S\left[3 - \cos\left(p_x a\right) - \cos\left(p_y a\right) - \cos\left(p_z a\right)\right]. \tag{195}$$

The spin waves of lowest energy occur at $\mathbf{p} = \mathbf{0}$, and correspond to a simultaneous rotation of all spins on the lattice by the same amount. (Obviously this does not cost any energy, since the relative orientations of all the spins remain the same.) For small wavevectors $|\mathbf{p}|a \ll 1$, we may perform a Taylor expansion of the cosines around $\mathbf{p} = \mathbf{0}$, and obtain

$$\varepsilon_{\mathbf{p}} \approx \hbar^2|J|Sa^2\mathbf{p}^2, \qquad |\mathbf{p}|a \ll 1. \tag{196}$$

Hence ferromagnetic spin waves have a quadratic dispersion law at low energies[57].

5.3.3 Checking the consistency of the spin wave approach

There is one last thing to do: namely, to check the consistency of our spin wave approach. This relies on the number of bosons on each site being small. But now that we have the energy spectrum (195), we can check this hypothesis explicitly.

[55] Literally, it is the energy of a boson in state $\mathbf{p}$. But recall that the bosons $\hat{a}_j$ represent excitations where a spin on site j is tilted away from the z axis—i.e. where S^z is reduced. So the bosons $\hat{a}_{\mathbf{p}}$ represent excitations where all of the spins are tilted away from their z axes, in a wavelike pattern described by wavevector $\mathbf{p}$. We call this kind of excitation a spin wave.

[56] If this sounds a bit like the way the action determines the particle trajectory (in classical mechanics) and the nature of quantum fluctuations (in quantum mechanics), that's because it is!

[57] Since this dispersion relation is the same as that for free particles, one might naïvely expect a $C_V \sim T$ specific heat as in the Fermi gas model. But this would be incorrect: the Fermi gas model deals with number-conserved fermions; ferromagnetic spin waves are both bosonic and not number-conserved. This highlights the *three* ingredients that go into thermodynamic power laws: dispersion relation; excitation statistics; and conservation laws.

The number of bosons on site j is given by

$$\hat{n}_j = \hat{a}_j^\dagger \hat{a}_j = \frac{1}{N} \sum_{\mathbf{p},\mathbf{q}} e^{i\mathbf{x}_j \cdot (\mathbf{q}-\mathbf{p})} \hat{a}_\mathbf{p}^\dagger \hat{a}_\mathbf{q}. \tag{197}$$

If the system is in equilibrium at a temperature T, we know that the average number of bosons in state $\mathbf{p}$ is given by the Bose-Einstein distribution function:[58]

$$\left\langle \hat{a}_\mathbf{p}^\dagger \hat{a}_\mathbf{p} \right\rangle = \frac{1}{e^{\beta\varepsilon_\mathbf{p}} - 1}, \tag{198}$$

where $\beta \equiv \frac{1}{k_B T}$ as usual. The cross-terms $\left\langle \hat{a}_\mathbf{p}^\dagger \hat{a}_\mathbf{q} \right\rangle$ vanish, since states of different wavevector are independent of each other. Hence the average number of bosons on site j is given by

$$\langle \hat{n}_j \rangle = \frac{1}{N} \sum_\mathbf{p} \frac{1}{e^{\beta\varepsilon_\mathbf{p}} - 1}; \tag{199}$$

this expression is independent of j, as one would expect for a translationally invariant system.

When N becomes large, we may convert this sum over $\mathbf{p}$ into an integral:

$$\langle \hat{n}_j \rangle = \frac{a^3}{8\pi^3} \int_{\text{BZ1}} \frac{1}{e^{\beta\varepsilon_\mathbf{p}} - 1} d^3\mathbf{p} \tag{200}$$

where the integral is taken over all values of $\mathbf{p}$ in the first Brillouin zone. The dominant contribution comes from states where the denominator is small, i.e. where $e^{\beta\varepsilon_\mathbf{p}} \approx 1$ — these are the low-energy states where $\varepsilon_\mathbf{p} \ll k_B T$. When the temperature T is small, we can therefore make two approximations without danger of serious error:

1. use the approximate form (196) for the spin wave spectrum;
2. extend the range of integration over all values of $\mathbf{p}$ (since by the time $\mathbf{p}$ reaches the Brillouin zone boundary the integrand is very small anyhow).

Hence

$$\begin{aligned} \langle \hat{n}_j \rangle &\approx \frac{a^3}{2\pi^2} \int_0^\infty \frac{1}{e^{\beta\hbar^2 S|J|a^2\mathbf{p}^2} - 1} p^2 dp \\ &= \frac{a^3}{4\pi^2} \left(\frac{k_B T}{\hbar^2 |J| S a^2} \right)^{3/2} \int_0^\infty \frac{\sqrt{y}}{e^y - 1} dy, \end{aligned} \tag{201}$$

where we have transformed into spherical polar co-ordinates (p, θ, ϕ), thus producing a measure $4\pi p^2 dp$ after the (trivial) angular integrations have been performed, and then used the substitution $y = p^2$ to attain the final line.

There are two things to note about the expression (201):

1. The integral on the right hand side converges. Hence the whole expression is $O(S^{-3/2})$, which is small when S becomes large. This means that the number of bosons on each site is small, and justifies our use of spin-wave theory. Conclusion: for the 3D Heisenberg ferromagnet, the ground state order is essentially classical, and the low-energy excitations are spin waves with a quadratic dispersion relation $\varepsilon_\mathbf{p} \propto \mathbf{p}^2$.

[58] Here, as in the black-body radiation case, the number of spin waves is not conserved and hence the chemical potential, μ, is zero.

2. The average number of bosons on each site goes as $T^{3/2}$. This fact could be used to calculate thermodynamic quantities such as the specific heat capacity.

An important feature of the calculation we have just performed is the measure factor $p^2 dp$ that came from the transformation into polar co-ordinates. In d dimensions, this would be $p^{d-1} dp$. Hence, for example, the integral in the two-dimensional case would read

$$\int_0^\infty \frac{1}{e^y - 1} dy, \tag{202}$$

and this integral is *not* convergent—the integrand has a non-integrable singularity at $y = 0$. The one-dimensional case is even worse. Hence, in one or two spatial dimensions, the number of bosons on each site *diverges* when calculated using spin wave theory. This simply means that spin wave theory is invalid in these cases, and an alternative approach must be sought.

This observation of the dimension-dependence of spin-wave theory calculations is related to a useful theorem in magnetism: the **Mermin-Wagner theorem**. It states, roughly, that

- No long-range magnetic order is possible in one dimension at all.
- No long-range order is possible in two dimensions, except at $T = 0$.

It is important to note that ferromagnets are an exception to this rule: it's obvious that, since the state $|\ldots \uparrow\uparrow\uparrow\uparrow \ldots\rangle$ is a strict ground state of the ferromagnet in any number of dimensions, one always has long-range order at $T = 0$ in this case[59].

5.4 Exercises

1. *Classical magnetism on the triangular lattice.* Determine the magnetic ordering pattern (or patterns) of the classical antiferromagnetic Heisenberg model on the two-dimensional triangular lattice.

2. *Antiferromagnetic spin-wave theory.* This question concerns the linearised spin-wave theory of the Heisenberg antiferromagnet on a cubic lattice.

   (a) Starting from the classical magnetic state of a Heisenberg antiferromagnet on a three-dimensional cubic lattice, write down the Holstein-Primakoff transformation for this case. (*Hint:* The classical magnetic order has two sublattices: the Holstein-Primakoff transformation will look different on each sublattice.)

   (b) Linearise this Holstein-Primakoff transformation, and thus obtain the momentum-space Hamiltonian for spin-waves in such an antiferromagnet. (*Hint:* You should find that this Hamiltonian, as well as number operators, contains anomalous terms of the sort $a^\dagger_{\mathbf{k}} b^\dagger_{-\mathbf{k}}$.)

   (c) Eliminate the anomalous terms from the spin-wave Hamiltonian via a **Bogoliubov transformation** of the type

$$\alpha_{\mathbf{k}} \equiv u_{\mathbf{k}} a_{\mathbf{k}} + v_{\mathbf{k}} b^\dagger_{-\mathbf{k}}, \tag{203}$$

$$\beta_{\mathbf{k}} \equiv u_{\mathbf{k}} b_{\mathbf{k}} + v_{\mathbf{k}} a^\dagger_{-\mathbf{k}}, \tag{204}$$

[59] The reason for this exception turns out to be that the order-parameter of a ferromagnet — the total magnetisation — is a conserved quantity, which makes quantum fluctuations ineffective in causing it to decay.

with a suitable choice of $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$, and thus obtain the dispersion relation for antiferromagnetic spin waves. (*Hint:* Make sure that your new operators $\alpha_{\mathbf{k}}$ and $\alpha_{\mathbf{k}}^\dagger$ still obey the fundamental bosonic commutation relations.)

(d) Comment on your result in comparison to the ferromagnetic case covered in the notes above.

6 A survey of experiment

6.1 In two sentences...

In this final lecture we shall study a couple of topical types of material in the light of the foregoing theoretical work. We shall present the basic ideas that high-temperature superconductivity arises from doping a Mott insulator, and that the heavy fermion compounds are described by a Kondo lattice model, without of course committing ourselves to either view.

6.2 The high-temperature superconductors: doping a Mott insulator?

6.2.1 Crystal structures

The so-called **high-temperature superconductors** began being discovered in 1986. Before then, it was widely believed (a) that all superconductors were described by the Bardeen-Cooper-Schrieffer (BCS) theory in which the superconductivity is driven by electron-phonon interactions, and (b) that the critical temperature, T_c, would never exceed 30 or 40 K.

The family of compounds that definitively destroyed the second of these beliefs, and severely damaged the first, is known as the **high-T_c superconductors**. Some examples of members of this family are $Bi_2Sr_2Ca_nCu_{n+1}O_{2n+6}$ (known as BSCCO), $YBa_2Cu_3O_{7-x}$ (known as YBCO), $La_{2-x}Sr_xCuO_4$ (known as LSCO), $Tl_2Ba_2CaCu_2O_8$, and $HgBa_2CaCu_2O_6$, though there are others. The crystal structures of a few of these are shown in Fig. 8. They all have a superconducting phase somewhere in their phase diagram, and the highest T_c in the family is now around 130 K, comfortably above the boiling point of nitrogen[60].

The unifying feature of all the compounds in this family is the CuO_2 layer(s) in their unit cell, for which reason they are often referred to as the **cuprates**. The number of such layers per unit cell varies from compound to compound, and this variation does seem to have an effect on T_c; nonetheless, most of the theoretical effort has focussed on high-T_c superconductivity as an essentially two-dimensional problem, with the physics going on in the copper oxide layers.

6.2.2 Basic experimental information

An important point about the cuprates is that the valence of the copper and oxygen ions can be controlled by altering the chemical constitution of the compound: this is known as **doping**. The undoped version of the compound (for example La_2CuO_4 in the LSCO family) is an antiferromagnetic insulator; doping causes this state to disappear and be replaced by a superconducting phase. The doping level $x = x_c$ at which T_c is maximal is referred to as **optimal doping**; compounds with $x < x_c$ are said to be **underdoped**, while those with $x > x_c$ are said to be overdoped. A putative generic phase diagram as a function of temperature (T) and doping (x) is shown in Fig. 9.

The superconducting phase appears to be fairly conventional in its nature, except for the fact that the symmetry of the order parameter is not the usual s-wave (same gap everywhere

[60] 77.36 K, in case you're curious.

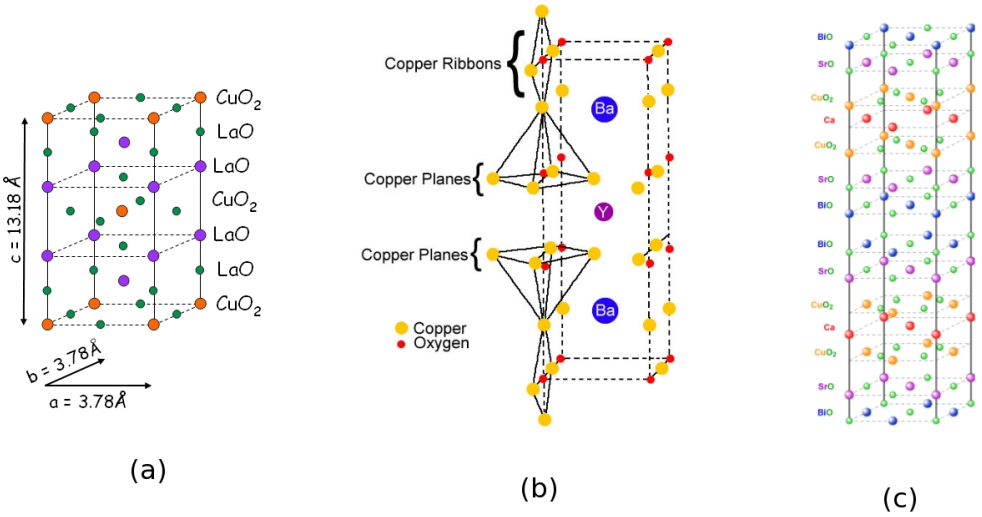


Figure 8: The crystal structures of three of the high-T_c materials: (a) La_2CuO_4; (b) $YBa_2Cu_3O_7$; (c) $Bi_2Sr_2Ca_nCu_{n+1}O_{2n+6}$.

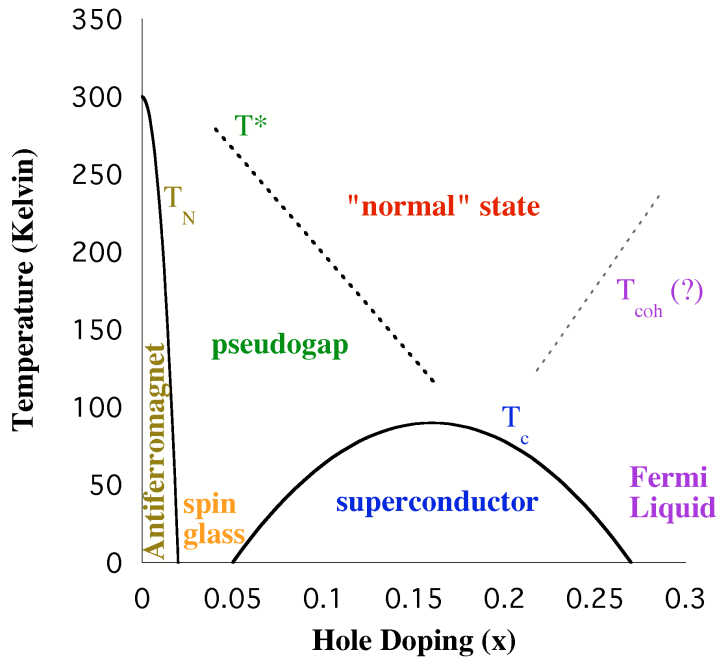


Figure 9: The 'generic phase diagram' of the cuprate superconductors. It's important to stress that no two phase diagrams from the cuprate family look exactly alike, so such pictures as this should be taken with a pinch of salt.

on the Fermi surface), but rather d-wave, which exhibits nodes at certain points in the Brillouin zone. What happens at $T > T_c$, on the other hand, is far from conventional. Firstly, there is a wide region of temperature in the underdoped compounds where a 'pseudogap' is observed in various excitation spectra of the system; part of this region is also characterised by very unusual thermoelectric properties. Secondly, at $T > T_c$ at optimal doping, the electrical resistivity is linear in T (as opposed to the quadratic dependence that Fermi liquid theory would predict, or the T^5 dependence expected from electron-phonon scattering). Also the magnetotransport properties in this unusual metallic state do not follow the power laws that one would expect from a semiclassical analysis. Even worse, photoemission experiments in this region do not show a quasiparticle peak in the spectral function, but rather a broad incoherent spectrum.

6.2.3 What should the model be?

The parent ($x = 0$) compound has one electron per unit cell, so the tight-binding model would predict metallic behaviour. The fact that it is actually insulating therefore suggests that we're dealing with a Mott insulator. The lattice is square, so there are no issues of frustration, and neutron scattering experiments confirm that the magnetic ordering pattern is just the Néel state that we would expect[61]. Furthermore, the spin-wave excitation spectrum fits the predictions of a Heisenberg model rather well. This, it seems, can all be understood in terms of the Hubbard/Heisenberg physics discussed in lectures four and five. It is this that led physicists in the late 1980s to converge (more or less) on the view that the high-T_c problem is essentially that of doping a Mott insulator.

What happens when the doping parameter x is changed[62]? Doping replaces La^{3+} by Sr^{2+}, and this causes an electron to be removed from (or equivalently a hole to be added to) one of the CuO_2 planes: hence the density of holes is x per unit cell. The active orbitals in the CuO_2 plane are the p_x and p_y orbitals on the oxygen ions, and the $d_{x^2-y^2}$ orbitals on the copper ions. This would suggest a **three-band Hubbard model** as a possible starting point.

However, an early argument from Zhang and Rice suggests that we can get away with a simpler model. The idea is that when an electron is removed from the plane, the resulting hole on the oxygen ion forms a singlet (known as a **Zhang-Rice singlet**) with the hole on the copper ion (in $3d^9$ electronic configuration), and thus we can think of the problem by considering only the unpaired electron (which carries spin $\frac{1}{2}$) living on each undoped copper site. That removes the oxygen p_x and p_y orbitals from the game, leaving only a **one-band Hubbard model** of the type we discussed in lecture four. By taking a $U \to \infty$ limit and strictly prohibiting double occupancy, we can reduce the size of the Hilbert space to three states per site[63]: $|0\rangle$, $|\uparrow\rangle$, and $|\downarrow\rangle$. The effective Hamiltonian in this reduced space is the **t-J model**:

$$H = -t \sum_{\langle ij \rangle \sigma} \left(c^\dagger_{i\sigma} c_{j\sigma} + c^\dagger_{j\sigma} c_{i\sigma} \right) + J \sum_{\langle ij \rangle} \left(\mathbf{S}_i \cdot \mathbf{S}_j - \frac{n_i n_j}{4} \right). \tag{205}$$

The first term describes the hopping of the holes (or equivalently electrons) from site to site,

[61] At this stage, you might legitimately worry about the Mermin-Wagner theorem. Doesn't it tell us that there can be no long-range magnetic order at $T \neq 0$ in two dimensions? Yes, it does; the get-out is to observe that the compound is of course not fully two-dimensional, so there will be some magnetic exchange coupling between neighbouring planes (call it $J_\perp$). A simple estimate of the Néel ordering temperature in such a situation gives $T_N \propto J_\parallel^{2/3} J_\perp^{1/3}$.

[62] We shall concentrate here, for simplicity, on the LSCO family, but the physics of doping is believed to be similar in the other cuprate compounds.

[63] Note that it's three, not two; because of the doping, there is no longer one electron per site, which means that some of them must be empty.

while the second represents the antiferromagnetic interaction between the spins of the electrons on the undoped sites.

It should be pointed out that neither the reduction from the three-band to the one-band Hubbard model, nor the further restriction to the $t - J$ model, are universally agreed upon in the community. There are also those who claim that, contrary to the motivating ideas of most of the past twenty years' work, inclusion of electron-phonon coupling is necessary to capture the physics.

6.2.4 Discussion

The first question to ask is whether the one-band Hubbard model, or the $t - J$ model (205), actually do exhibit superconductivity in some range of doping. This question has proved surprisingly difficult to answer: different analytic approximations give conflicting answers, and I think it's fair to say that the numerical evidence remains inconclusive[64].

Assuming that the model does require some modification, there are several things that could be added, and opinion differs on which of these ingredients are essential. Here are a few, with brief comments:

- *Disorder.* The chemical doping, as well as changing the carrier density, introduces significant disorder into the compound. As tunnelling experiments on BSCCO show, this is reflected in a fairly broad distribution of values of the superconducting gap Δ as a function of position, including (at low dopings) some regions where there appears to be no superconductivity at all. To what extent disorder affects the superconducting behaviour, or indeed the unusual normal state properties, remains an open question.

- *Further-neighbour hopping.* The simplest versions of the Hubbard and $t - J$ Hamiltonians feature hopping between nearest neighbour lattice sites only. However, the neutron-scattering measurements on the parent compound of the LSCO family do not fit the spin-wave spectrum derived from the nearest-neighbour Heisenberg model; instead, some amount of further-neighbour hopping (and ring exchange: correlated hopping of several holes around a plaquette of the lattice) must be included to fit the data. How important this further-neighbour hopping is for superconductivity is unknown.

- *Further-neighbour Coulomb repulsion.* In the same way that the assumption of nearest-neighbour hopping is approximate, the assumption that there is Coulomb repulsion only between holes on the same site is rather a crude approximation. It might be all right in a good metal, since there the Coulomb interaction is screened by charge displacements relatively quickly; but in a compound that is close to becoming an insulator, the approximation of an on-site U may be less reliable. Nearest-neighbour, or even long-range, Coulomb repulsion has been proposed by several researchers as the missing ingredient in understanding cuprate physics.

- *Phonons.* Over the last few years the community has begun retreating somewhat from its belief that phonons were irrelevant for explaining high-T_c superconductivity. The original reason for thinking this was that the isotope effect in these compounds (the change in

[64] One of the problems of doing numerical simulations of Hubbard or $t-J$ models is the sheer size of the Hilbert space. With four states per site (in the Hubbard model case), there are 4^N states for N sites: this means that to exactly diagonalise a five-by-five section of the square lattice requires a Hilbert space of dimension $4^{25} \approx 10^{15}$. It's not quite that bad, since conservation laws permit the space to be subdivided somewhat, but it still means that telling the genuine physics of the bulk model from finite-size effects is not so easy.

Left: Tetragonal unit cells of the $ThCr_2Si_2$ and CeM_2X_2 (Ce122) structure where M = Cu, Ni, Ru, Rh, Pd, Au, .. (black circles); X = Si, Ge (small grey circles) and Ce (large grey circles). Right: Conventional unit cell of the $CeMIn_5$ (Ce115) structure whith M = Co, Ir, Rh (black circles); In (small grey circles) and Ce (large grey circles).

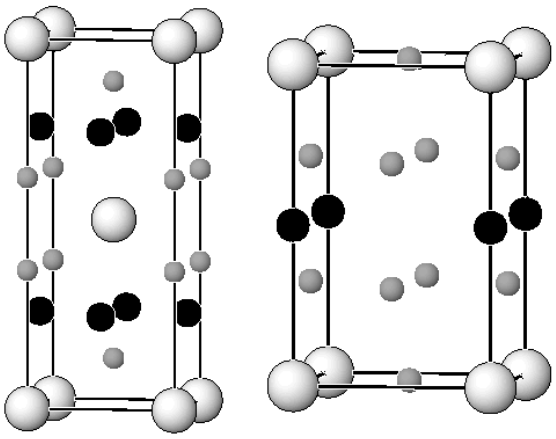


Figure 10: The structure of some of the heavy fermion materials. Notice that they all contain a lanthanide or actinide element, suggesting that f-electron physics is important. Figure courtesy of Frank Steglich.

T_c when ^{16}O is replaced by ^{18}O) is very small; but recently several pieces of evidence have surfaced — mainly from photoemission and tunnelling studies — that suggest that a particular phonon mode due to the oxygen ions is quite strongly coupled to the electronic sector. Hybrid theories, such as 'gossamer superconductivity' (in which the phonons are required to kick-start superconductivity, but T_c is so high because they are working on an electronic state that is already very strongly susceptible to electron-electron pairing), have also appeared.

Let's close this section on an optimistic note. The initial experimental observation of superconductivity was made by Kamerlingh Onnes in 1911; Bardeen, Cooper, and Schrieffer published their microscopic theory of superconductivity in 1957. Therefore, assuming that the high-T_c problem is at least as hard as the 'low-T_c' one, we need not feel embarrassed at our lack of a solution until about 2032…

6.3 Heavy fermion materials: the Kondo lattice model?

6.3.1 Crystal structures

The **heavy-fermion compounds** have been mentioned already in these lectures. They began to be discovered in 1979, and the crystal structures of some of them are shown in Fig. 10. Notice that, unlike the cuprates, they are not (or at least not always) quasi-two-dimensional, and indeed some are fully cubic. It is important to note that they all contain a lanthanide or actinide element, be it uranium in UPt_3, ytterbium in $YbRh_2Si_2$, or cerium in $CePd_2Si_2$. This suggests that the physics of f-electrons is important here, as we shall discuss in the following.

6.3.2 Basic experimental information

The heavy fermion compounds get their name because of the extremely large effective masses observed in their metallic states: $\frac{m^*}{m}$ (where m^* is the effective mass and m the bare electron mass) can get up to values of order 1000, though typically it is more like 200 or 300. Still, this is a huge mass-enhancement compared to that observed in most Fermi liquids, and thus

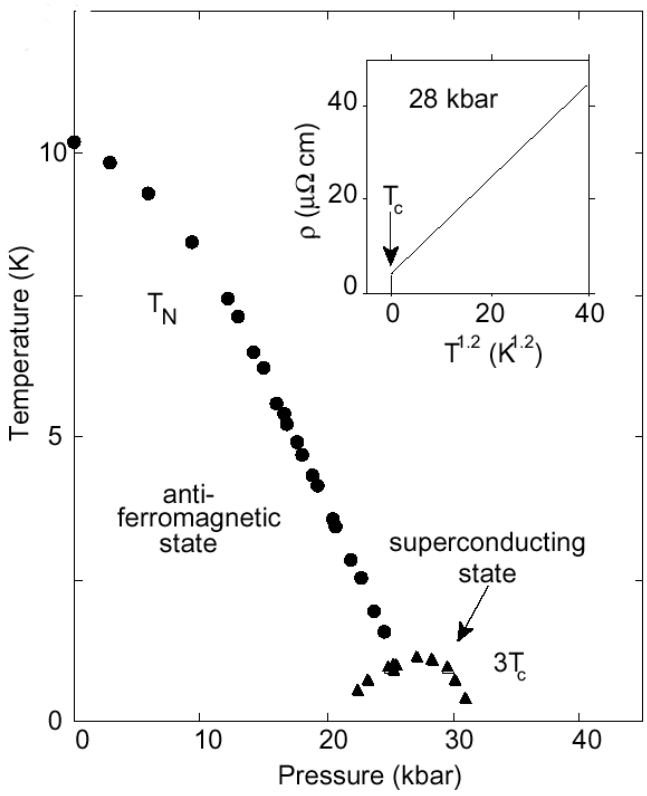


Figure 11: The phase diagram of $CePd_2Si_2$. The phase on the left is an antiferromagnetic metal, whose Néel temperature is forced to zero by increasing the pressure. Around the point where T_N goes to zero, a bubble of superconductivity forms. At higher pressures the compound exhibits Fermi liquid behaviour, but with a dramatically enhanced effective mass. Figure from N.D. Mathur *et al.*, *Nature*, 1998.

these materials are described as having a **heavy Fermi liquid** phase. This phase, however, is not alone in the phase diagram; it is usually accompanied by a magnetic phase (which may be antiferromagnetic or ferromagnetic), and sometimes by a 'bubble' of superconductivity around the point where the magnetic phase disappears. A typical phase diagram (for $CePd_2Si_2$) is shown in Fig. 11. Unlike in the cuprates, where chemical doping is the only way of exploring the different phases, in the heavy fermion materials this can also be achieved with pressure, which has the advantage that no extra disorder is introduced into the system as one moves from the magnetic, through the superconducting, to the heavy Fermi liquid state. Notice the inset: as in the cuprates, an anomalous power law in the resistivity ($\rho \sim T^{1.2}$) is seen above the superconducting dome.

6.3.3 The Doniach phase diagram and the Kondo lattice model

A first stab at the physics of the heavy fermion compounds starts from the observation (mentioned above) that there are f-electron elements in the chemical formula. f-orbitals are fairly narrow compared to s- or p-orbitals, which suggests that the energy cost for double occupancy should be higher in the f-orbital case. Consequently, if we model the s-electron band(s) by one Hubbard model, and the f-electron bands by another, it is plausible that the s-electron band(s)

could be in the metallic state (due to their relatively small U/t ratio), while the f-electron band(s) could be in the Mott insulating state (since their U/t is larger). That in turn suggests that we might consider the f-electrons as local moments, and therefore represent them by spin operators, as we did above in the derivation of the Heisenberg model. In the simplest case these would be spin-$\frac{1}{2}$ operators: this is appropriate for the cerium compounds (in which the cerium ions have a single f-electron) and for the ytterbium compounds (in which the ytterbium ions have a single f-hole). The case of the uranium compounds, with three f-electrons, is more complicated. But sticking to spin-$\frac{1}{2}$ for the moment, this suggests the following model:

$$H = -t \sum_{\langle ij \rangle \sigma} \left(c^\dagger_{i\sigma} c_{j\sigma} + c^\dagger_{j\sigma} c_{i\sigma} \right) + \frac{J}{2} \sum_{i\sigma\tau a} c^\dagger_{i\sigma} \sigma^a_{\sigma\tau} c_{i\tau} S^a_i, \tag{206}$$

where $c^\dagger_{i\sigma}$ creates a conduction electron on lattice site i with spin σ, and S^a_i is a vector of spin-$\frac{1}{2}$ operators representing the localised f-electron on site i. This is called the **Kondo lattice model**, since it is a version of the Kondo model (98) with an impurity at every lattice site rather than just one at the origin.

How to solve this model is not known: it is much harder than the single-impurity Kondo problem because of the possibility of coherent multiple scattering of conduction electrons by different impurity spins. But an interesting simple argument about its physics was advanced by Doniach; it is based on the idea that the physics of (206) is governed by the competition between two effects:

- *Kondo screening.* Recall that, in the single-impurity Kondo problem, the spin of the impurity becomes screened below a characteristic temperature scale

$$T_K = D \exp\left(-\frac{1}{2\nu_0 J} \right), \tag{207}$$

  where D is the bandwidth and ν_0 is the density of states at the Fermi level. This screening leads to a restoration of Fermi liquid behaviour at temperatures $T \ll T_K$. Let us assume that this tendency persists in the Kondo lattice model. It should therefore lead to a Fermi liquid phase where the magnetic susceptibility is constant.

- *Ruderman-Kittel-Kasuya-Yosida (RKKY) interactions.* The competing effect, which is not present in the single-impurity model, is that the impurity spins can 'talk' to each other via the conduction electrons. Essentially the way this works is that one impurity excites Friedel oscillations in the Fermi sea, thus changing the magnetic field felt by another impurity, and leading to an effective interaction $J_{\rm RKKY} \mathbf{S}_1 \cdot \mathbf{S}_2$. The coefficient $J_{\rm RKKY}$ depends on the distance between the impurities, and oscillates from ferromagnetic to antiferromagnetic and back again as the impurities get further apart[65]. This interaction is called the **RKKY interaction**, after its discoverers. Since it involves two actions of the J-term in (206) — one for $\mathbf{S}_1$ to produce the Friedel oscillations, and another for $\mathbf{S}_2$ to detect them — we must have

$$J_{\rm RKKY} = AJ^2, \tag{208}$$

  where A is a constant that could be calculated. This RKKY interaction therefore tends to produce long-range magnetic order by influencing the alignment of neighbouring spins; this long-range order will be antiferromagnetic or ferromagnetic depending on the sign of $J_{\rm RKKY}$.

[65] This is a simple consequence of the fact that the Friedel oscillations themselves change sign with a wavelength λ_F.

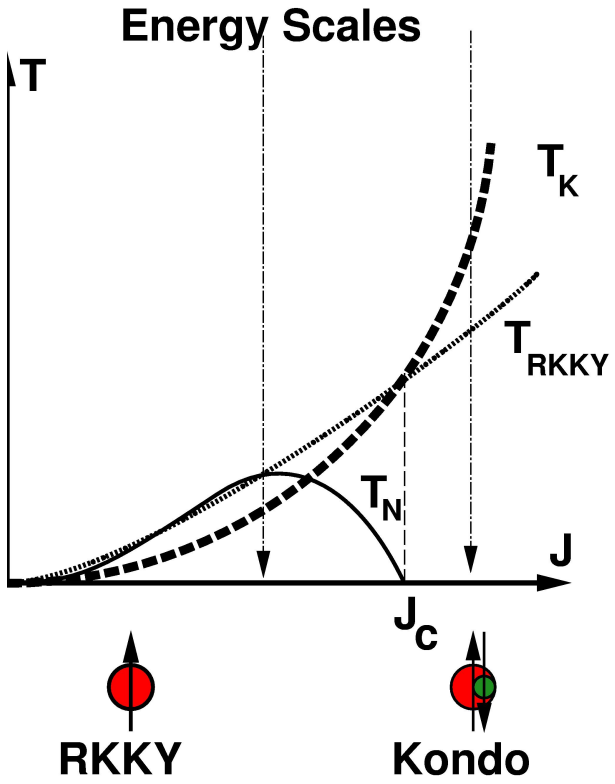


Figure 12: The Doniach phase diagram, plotting the J-dependence of the energy scales (207) and (208). For small values of J, $T_{\mathrm{RKKY}} > T_K$, and thus a magnetically ordered state forms; for large values of J, $T_K > T_{\mathrm{RKKY}}$, leading to a heavy Fermi liquid. The point at which the two states become degenerate is known as a quantum critical point (QCP). Figure courtesy of Antonio Castro Neto.

The idea is that applying pressure changes the value of J (by changing the lattice spacing and therefore changing the quantum mechanical overlap integrals from which J is derived). This affects both T_K and J_{RKKY}, but in different ways because of their different functional dependences on J. When J is small (at low pressure), the RKKY interaction wins and a magnetically ordered metallic phase results. When J is large (at high pressure), the Kondo screening wins and each impurity is screened by the conduction sea, resulting in the heavy Fermi liquid phase. A sketch of the J-dependences of T_K and J_{RKKY} is shown in Fig. 12. It is often referred to as the **Doniach phase diagram**.

The critical value J_c (which corresponds to a critical value of pressure p_c) where the transition from one state to the other occurs is called a **quantum critical point** (QCP). It is so called because, whereas at ordinary critical points a transition from one thermodynamic state to another occurs as a function of temperature, at a QCP a transition from one quantum ground state to another occurs as a function of an external tuning parameter (in this case the pressure). It is the quantum fluctuations associated with this QCP that are believed to be responsible for the unusual resistivity and other transport signatures in the region above the QCP in the phase diagram.

6.3.4 Discussion

There are several points to make about the foregoing analysis:

- It is not quantitatively accurate. If one makes a reasonable estimate of the numbers that ought to go into (207) and (208), one finds that J_c comes out to be a very high energy, and hence one would really expect all of these compounds to be magnetic at all pressures. This failure is doubtless due to the fact that we have naïvely transcribed the single-impurity Kondo scale and assumed that it applies unaltered to the multi-impurity problem, whereas surely the true physics is much more complicated than this.

- The Doniach phase diagram is a good qualitative explanation of why magnetism gives way to a heavy Fermi liquid, but it does not provide a way of thinking about why superconductivity is sometimes seen in the region around the QCP. The explanation for this is widely believed to be the formation of Cooper pairs due to magnetically mediated (rather than phonon-mediated) pairing: near the QCP the magnetic fluctuations are very strong, since magnetic order is about to set in, and these are believed to act as a pairing mechanism. There are some theories of how this works, but they are rather phenomenological in character and much work remains to be done on this aspect of the problem.

- Another interesting question concerns the Fermi surface volume. In the magnetic phase, the conduction sea is made only of the s-electrons, whereas in the heavy Fermi liquid phase it is made of the s- and f-electrons hybridised together via the Kondo effect. This appears to imply that the Fermi surface should have a smaller volume in the magnetic state than in the heavy Fermi liquid state, since it contains fewer electrons in the former case. That in turn leads to the question of how the transition from one Fermi surface volume to the other occurs. Gradually, as J is increased? Suddenly, as a transition at $J = J_c$? This question remains open, and is the subject of ongoing theoretical work; recent proposals have also been made to detect the Fermi surface volume more-or-less directly using magnetotransport measurements.

- Finally, it would be remiss to leave this topic without saying something about the relationship between the cuprate superconductors and the heavy fermion compounds. Certainly at a glance the phase diagrams in Figs. 9 and 11 look somewhat similar: they both have a dome of superconductivity close to the end of a magnetic transition line, they both have a Fermi liquid beyond the superconducting state[66], and they both exhibit anomalous transport properties in the region above the superconducting dome. The ideas that therefore (a) the Cooper pairing in the cuprates is caused by magnetic fluctuations, and (b) that there is a QCP buried under the cuprates' superconducting dome, have gained some currency. Others, however, stress the contrasts between the two cases: for example, in the cuprates the magnetic order vanishes before the superconductivity begins, which means that if there is a QCP under the superconducting dome it does not seem to be directly connected with the line of magnetic transitions $T_N(x)$. One could counter that the appropriate line to look at is in fact T^*, the pseudogap temperature; but there is fairly good evidence that this line does not correspond to a thermodynamic transition. And, as mentioned earlier, a view that phonons are in fact of some importance in the cuprates is beginning to emerge, and this would draw another line between them and the heavy fermion compounds.

[66] This is not completely established in the case of the cuprates, but recent evidence from angular magnetoresistance oscillations makes a plausible case for it.

STM

Statistical Mechanics

Richard Blythe

University of Edinburgh

Physics by the Lake 2018

Cumberland Lodge, The Great Park, Windsor

Preface

Statistical mechanics is the study of unpredictability in physics. Most often we are concerned with the apparent randomness of particle motion in systems that are thermally agitated by some external reservoir whose degrees of freedom we don't care too much about. In this course, we will examine the macroscopic consequences of this microscopic stochastic dynamics: equations of state, critical phenomena, Brownian motion, fluctuation-dissipation relations and far-from-equilibrium work theorems.

A difficulty in preparing this course is that the students come with wildly varying backgrounds. Some will have the dimmest recollection of a Boltzmann factor and its role in condensed matter physics; others will be using statistical mechanics daily in their PhD research. I learnt a lot about the subject in preparing these lectures; I hope that those 'professionals' who find many parts of the material familiar will do so too.

To help us all along, some sections containing more advanced material or lengthy derivations are indicated with an asterisk $*$. The 'professionals' should pay particular attention to these sections; meanwhile, those for whom this course is less familiar may wish to skim over these sections on a first encounter. Additionally, there are three tutorial exercises associated with each of the six lectures. The grading 'A', 'B' and 'C' indicates successively more involved calculations rather than an increase in complexity or obscurity of the underlying concepts. Whilst the 'C' questions are emphatically recommended to the 'professionals', they could be omitted by those wanting to cover the full breadth, rather than the depth, of the material.

There is no one book that I am aware of that covers this course. A gentle introduction to phase transitions—the focus of Lectures 2 and 3—is by Yeomans (1992) (see next page for the bibliography). An excellent book that gets into the real details of the renormalisation group (touched on in Lecture 3) is Goldenfeld (1992). The rest of the material concerns dynamics, and unfortunately I'm yet to find a nice treatment of this subject—and as it's a swiftly-moving field it will be a while before we get something definitive. In the meantime, I'd suggest Reichl (2009) for a more classical treatment, or Evans and Morriss (2008) for one that, despite being skewed towards molecular dynamics simulations, contains some good material on more modern theories, such as fluctuation relations and thermostats (a topic I've not delved into here, but is relevant to the course). For stochastic dynamics, many eminent people swear by van Kampen (1992), whilst others swear at it.

I thank previous lecturers of this course, particularly Martin Evans and Mike Swift, for providing me with fine examples to follow, and Lawrence Mitchell for showing me how to change the font in LaTeX.
Despite the fact that I have given this course several times now, and it is apparently converging to a stable limit, myriad errors doubtless remain in the notes. Do point them out to me so that we may achieve more rapid relaxation to equilibrium.

Richard A Blythe — Edinburgh, June 29, 2018
R.A.Blythe@ed.ac.uk

References cited in these notes

A J Bray. Theory of phase-ordering kinetics. *Adv. Phys.*, 43:357, 1994.

A D Bruce and N B Wilding. Scaling fields and universality of the liquid-gas critical point. *Phys. Rev. Lett.*, 68:193, 1992.

C Bustamante, J Liphardt, and F Ritort. The nonequilibrium thermodynamics of small systems. *Physics Today*, 58:43, 2005.

G E Crooks. Path-ensemble averages in systems driven far from equilibrium. *Phys. Rev. E*, 61:2361, 2000.

D J Evans and G P Morriss. *Statistical mechanics of nonequilibrium liquids*. Cambridge University Press, Cambridge, 2008.

D J Evans, E G D Cohen, and G P Morriss. Probability of second law violations in shearing steady states. *Phys. Rev. Lett.*, 71:2401, 1993.

N Goldenfeld. *Lectures on phase transitions and the renormalization group*. Addison-Wesley, Reading, MA, 1992.

P C Hohenberg and B I Halperin. Theory of dynamic critical phenomena. *Rev. Mod. Phys*, 49:435, 1977.

S Jabbari-Farouji, D Mizuno, M Atakhorrami, F C MacKintosh, C F Schmidt, E Eiser, G H Wegdam, and D Bonn. Fluctuation-dissipation theorem in an aging colloidal glass. *Phys. Rev. Lett.*, 98:108302, 2007.

C Jarzynski. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78:2690, 1997.

J Liphardt, S Dumont, S B Smith, I Tinoco, and C Bustamante. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science*, 296:1832, 2002.

M E J Newman and G T Barkema. *Monte Carlo methods in statistical physics*. Clarendon Press, Oxford, 1999.

L Onsager. Reciproal relations in irreversible processes. I. *Phys. Rev.*, 37:405, 1931a.

L Onsager. Reciproal relations in irreversible processes. II. *Phys. Rev.*, 38:2265, 1931b.

P Pradhan, Y Kafri, and D Levine. Nonequilibrium fluctuation theorems in the presence of local heating. *Phys. Rev. E*, 77:041129, 2008.

L E Reichl. *Modern course in statistical physics*. Wiley-VCH, London, 3 edition, 2009.

H Risken. *The Fokker-Planck equation: Methods of solution and applications*. Springer-Verlag, Berlin, 1989.

N G van Kampen. *Stochastic processes in physics and chemistry*. North-Holland, Amsterdam, 1992.

G Volpe, L Helden, T Brettschneider, Jan Wehr, and C Bechinger. Influence of noise on force measurements. *Phys. Rev. Lett.*, 104:170602, 2010.

G M Wang, E M Sevick, E Mittag, D J Searles, and D J Evans. Experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales. *Phys. Rev. Lett.*, 89: 050601, 2002.

H S Wilf. *Generatingfunctionology*. A K Peters, Wellesley, MA, 2006.

U Wolff. Collective Monte Carlo updating for spin systems. *Phys. Rev. Lett.*, 62:361, 1989.

J M Yeomans. *Statistical mechanics of phase transitions*. Clarendon Press, Oxford, 1992.

Lecture 1

The zeroth law of thermodynamics

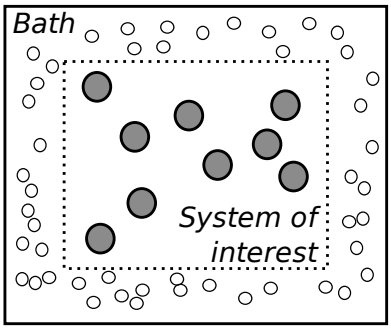
The impressive thing about about a system of interest coupled to a thermal environment is that when equilibrium is reached, its macroscopic properties do not depend on the microscopic structure of the environment or the details of the coupling. Instead, only a small number of macroscopic parameters, such as the temperature, matter. This is essentially the content of the zeroth law of thermodynamics. In this lecture, we show how the environment can be 'integrated out' to provide a self-contained description of both statics and dynamics of the system of interest at the microscopic and macroscopic scales. We frame the discussion with two concrete examples.

1.1 Exhibit A: a Lennard-Jones fluid

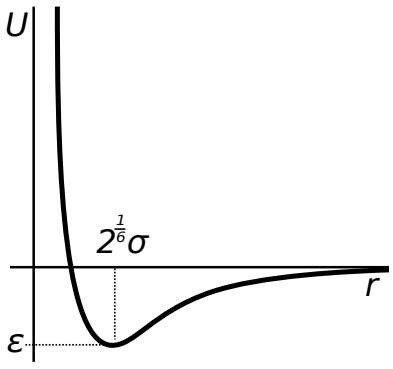
A popular model for the interaction energy of a pair of atoms separated by a distance r is the **Lennard-Jones 6-12** potential:

$$U_{\mathrm{LJ}}(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{1.1}$$

where ϵ and σ are parameters characterising the depth and range of the potential—see Fig. 1.1. The attractive r^{-6} piece can be argued for by considering the interaction between two electric dipoles. The repulsive r^{-12} part is purely phenomenonological.

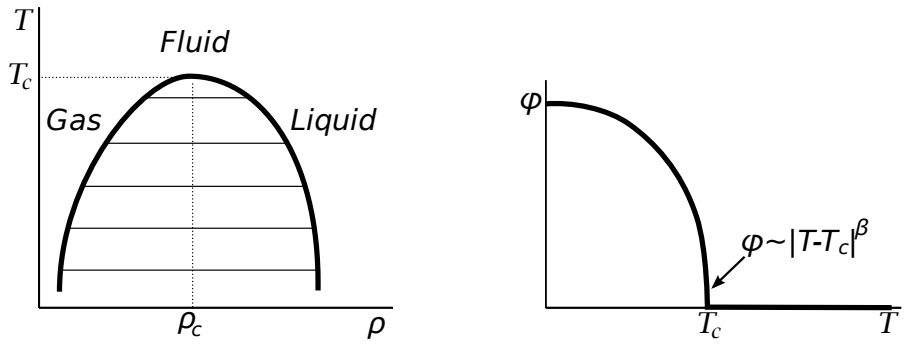


(a) *A combined system comprising the fluid particles of the 'system of interest' and those of the 'bath'*



(b) *The Lennard-Jones 6–12 potential*

Figure 1.1 *The Lennard-Jones fluid system*



(a) *Phase diagram* **(b)** *Order parameter* $\phi = \rho_\ell - \rho_g$

Figure 1.2 *Macroscopic phase behaviour of the Lennard-Jones fluid system*

The Hamiltonian for the combined system (i.e., system of interest + bath) is

$$H(\mathbf{x},\mathbf{p}) = \sum_i \frac{p_i^2}{2m_i} + \sum_{\langle i,j\rangle} U_{\mathrm{LJ}}(|x_i - x_j|) + \sum_i U_{\mathrm{ext}}(x_i) \tag{1.2}$$

where x_i and p_i are individual (possibly vectorial) particle positions and momenta; m_i their masses; $\langle i,j\rangle$ represents distinct particle pairs; bold-face $\mathbf{x}$ and $\mathbf{p}$ represent the state of the whole system; and $U_{\mathrm{ext}}(x)$ is some external potential that, for example, confines all particles to a box (as in Fig. 1.1a). The microscopic equations of motion are then given as usual by Hamilton's equations:

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = -\frac{\partial H}{\partial x_i} \quad \text{and} \quad \frac{\mathrm{d}x_i}{\mathrm{d}t} = \frac{\partial H}{\partial p_i}\ . \tag{1.3}$$

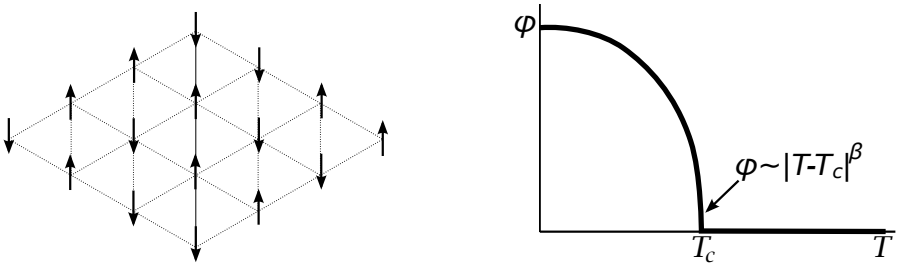
Macroscopically, the Lennard-Jones system is known to exhibit a number of phase transitions. We will examine the fluid-fluid (i.e., liquid-gas transition) in its standard representation as a phase diagram in the density-temperature (ρ–T) plane, shown in Fig. 1.2a. This shows, for any given combination of ρ, T the state of the system that is reached at equilibrium. Most interesting is the presence of a critical temperature T_c, above which one sees only one fluid phase, and below which there is a transition from a low-density gas phase to a high-density liquid phase with a coexistence region at intermediate densities.

This phase transition is conventionally characterised by means of an **order parameter**, denoted ϕ. Order parameters are chosen such that they vanish in a disordered phase, and take on a nonzero (normally positive) value indicating the degree of order present in the ordered phase. For the fluid system, the order parameter is set to zero in the absence of phase separation (i.e., above T_c) and is set to the difference between the liquid and gas densities when the fluid phase separates, *i.e.*, $\phi = \rho_\ell - \rho_g$. If the temperature is lowered below T_c at the critical density ρ_c, the order parameter grows from zero as $\rho \sim |T - T_c|^{\beta}$, as shown in Fig. 1.2b. For a two-dimensional Lennard-Jones fluid, the **critical exponent** β has the value $\frac{1}{8}$.

1.2 Exhibit B: An Ising magnet

Two magnetic moments in a solid have an interaction energy $\Delta E_{ij} = -J_{ij}\sigma_i \cdot \sigma_j$ where, in principle, σ_i and σ_j are quantum-mechanical spins whose components are represented mathematically by Pauli matrices and the exchange J_{ij} is a consequence of the antisymmetry of fermionic wavefunctions.

It turns out that a great deal of insight can be obtained from a purely classical simplification of the quantum mechanics. We take $J_{ij} = J$ constant for nearest neighbours (and zero otherwise), and uniaxial



(a) *The Ising Model on the triangular lattice*

(b) *Order parameter* $\phi = M$

Figure 1.3 *The Ising model: a magnetic system comprising uniaxial 'spins'*

scalar spins $\sigma_i \to s_i = \pm 1$. The resulting model is called the **Ising model**, and has an internal energy function

$$U_{\text{Ising}}(\mathbf{s}) = -J \sum_{\langle ij \rangle} s_i s_j \tag{1.4}$$

for the degrees of freedom in the system of interest. See Fig. 1.3a. Given its origins in quantum mechanics and subsequent stylisation, one does not normally write down a dynamics analogous to (1.3) which describes the time evolution of the system and energy exchanges with the environment. As we will see later in this lecture, we derive dynamics by appealing to general physical principles.

The significance of the Ising model is that, in two or more dimensions, it exhibits a phase transition. Above a critical temperature T_c, the magnet exhibits zero magnetisation M, a quantity defined as

$$M = \sum_i \langle s_i \rangle \ , \tag{1.5}$$

where $\langle \cdot \rangle$ is an average over the thermal **noise** in the environment. Below T_c, the magnet spontaneously orders, i.e., makes a transition to a nonzero magnetisation M (indicating that a more spins point one way than the other). It is clear that $\phi = M$ is a suitable order parameter for the transition, and exhibits the behaviour shown in Fig. 1.3b. The fact that Figs. 1.3b and 1.2b are identical is not a typesetting error. The behaviour of the order parameter near the critical temperature T_c is not only qualitatively similar in both cases, but quantitatively similar. Just below the critical temperature, $\phi \sim |T - T_c|^\beta$ with $\beta = \frac{1}{8}$ for *both* the Lennard-Jones fluid and the nearest-neighbour Ising model in two dimensions (although the prefactor will be different). This quantitative similarity is called **universality** and is closely associated with **critical phenomena**.

1.3 Free energy

Phase diagrams are a useful way to visualise the equilibrium behaviour of a statistical mechanical system. How do we determine equilibrium state variables like pressure or magnetisation that allow the different phases to be identified? If we were dealing with an athermal system (such as balls connected by springs), we would ask for all forces to balance—or equivalently, the minimum a potential energy function: $\nabla U(\mathbf{x}) = 0$.

Because the system of interest can exchange energy with the surrounding bath, we seek not the minimum of the potential energy, but a **free energy** that takes into account the coupling to the bath. It also takes into account—via an **entropy** term—that instead of equilibrium being a single configuration in which all positions remain fixed, it is in fact a **statistical ensemble** of configurations in which all particles keep moving and colliding with one another. The equilibrium state is one in which the *macroscopic* (average) properties, such as the energy or pressure, remain constant.

1.4 Fluid systems: Helmholtz free energy

When the volume of a system of interest is fixed—as is often the case for a fluid—the appropriate form of the free energy is the **Helmholtz free energy**. Suppose that the internal energy of the fluid is $\tilde{U}$. Then, the Helmholtz free energy of that system, when coupled to a heat bath with temperature T, is

$$\tilde{F}(\tilde{U}) = \tilde{U} - T\tilde{S}(\tilde{U}) \tag{1.6}$$

where $\tilde{S}(\tilde{U})$ is the entropy associated with a macrostate of energy $\tilde{U}$. Why all the twiddles? This is not the equilibrium free energy. This is obtained by minimising the twiddled free energy with respect to $\tilde{U}$ (or, indeed, any other unconstrained quantity). Thus,

$$\frac{\mathrm{d}}{\mathrm{d}\tilde{U}}\tilde{F}(\tilde{U})\bigg|_{\tilde{U}=U} = 0 \tag{1.7}$$

defines the equilibrium energy U and therewith the equilibrium entropy $S = \tilde{S}(U)$ and free energy $F = U - TS$. Note that the equilibrium free energy is rarely distinguished from the function that needs to be minimised to find the equilibrium. This can lead to confusion.

Once we have the free energy, we have everything else via derivatives. Specifically, for a fluid system constrained to have volume V and temperature T we find the entropy and pressure via

$$\boxed{S = -\frac{\partial F(T,V)}{\partial T} \quad \text{and} \quad p = -\frac{\partial F(T,V)}{\partial V}}\,. \tag{1.8}$$

1.5 Magnetic systems: Gibbs free energy

If one magnetic system is in contact with another, it is typically the case that the magnetisation M of each may fluctuate. Equilibrium is then reached when the field h due to one magnet equals that due to the other. The appropriate free energy is the **Gibbs free energy**

$$\tilde{G}(T,h) = \tilde{U} - T\tilde{S} - h\tilde{M} \tag{1.9}$$

where, again, we have placed tildes on the unconstrained quantities. Again, equilibrium is reached by minimising with respect to these. Once the equilibrium free energy has been identified, the entropy and magnetisation are given by

$$\boxed{S = -\frac{\partial G(T,h)}{\partial T} \quad \text{and} \quad M = -\frac{\partial G(T,h)}{\partial h}}\,. \tag{1.10}$$

If it is the magnetisation of the system that is fixed, then one would instead use the Helmholtz free energy $F(T,M)$. One can transform from the Helmholtz to the Gibbs free energy by first differentiating to obtain the field as a function of temperature and magnetisation,

$$h(T,M) = \frac{\partial F(T,M)}{\partial M}\,, \tag{1.11}$$

inverting to get $M(T,h)$. Then, one uses the relationship

$$G(T,h) = F(T,M[T,h]) - hM(T,h)\,. \tag{1.12}$$

Exercise 1A revises the connection between the different free energies if you have forgotten it.

1.6 Microscopic dynamics at equilibrium: detailed balance

Free energy minimisation is a macroscopic principle for determining the equilibrium state of a system coupled to a bath. The corresponding microscopic principle involves rules for transitions between microstates and is known as **detailed balance**.

Consider, for example, a fluid system governed by Hamilton's equations (1.3). If we cover up the fluid particles in the 'bath' portion of the system, the motion of the particles of interest becomes apparently stochastic due to interactions with the invisible bath particles. The microscopic dynamics must therefore specified in terms of a set of **transition probabilities** between different microstates.

Let us illustrate with a spin system which, at some time t, has the microscopic configuration $\mathbf{s}$. After some time interval δt, it has some probability of being in a configuration $\mathbf{s}'$ *given* that it was in state $\mathbf{s}$ at time t. We denote this probability $P(\mathbf{s} \to \mathbf{s}')$. If the system of interest is at equilibrium with the bath, then all configurations in the combined system are equally likely. This fact, combined with the property of **time-reversal symmetry** in the underlying dynamics leads to the following relationship between pairs of transition probabilities:

$$\frac{P(\mathbf{s} \to \mathbf{s}')}{P(\mathbf{s}' \to \mathbf{s})} = \exp\left(\frac{\Delta Q}{kT}\right) . \tag{1.13}$$

This is the detailed balance relation for spin systems. The quantity ΔQ is the heat transferred *from* the system of interest *into* the heat bath in the transition $\mathbf{s} \to \mathbf{s}'$. We will derive this relation from the underlying physical principles at the end of the lecture: in the meantime, we shall explore its consequences.

A key point is that the energy of the combined system is conserved. Furthermore, the heat surrendered to the bath in the transition $\mathbf{s} \to \mathbf{s}'$ must be reabsorbed in the reverse transition $\mathbf{s}' \to \mathbf{s}$. Thus, transitions that involve an increase in the entropy of the bath (or, equivalently a lowering in energy of the system of interest) are favoured by the dynamics.

1.7 Generating equilibrium phase-space trajectories: the Monte Carlo method

The detailed balance relation (1.13) provides a means to generate a **trajectory** through the space of microstates of the system of interest that samples all states of the *combined* system uniformly *without* the need to directly simulate the degrees of freedom of the bath. What one does in practice is to construct a set of transition probabilities that satisfy (1.13), then, given a configuration $\mathbf{s}$ at time t, one chooses the next configuration to be $\mathbf{s}'$ with probability $P(\mathbf{s} \to \mathbf{s}')$. Iterating this procedure then generates a sequence of configurations. The number of times that configuration $\mathbf{s}$ is generated is proportional to its equilibrium probability: thus one can measure equilibrium properties of the system by averaging over this sequence of configurations.

We now explain these steps in more detail with reference to a specific algorithm:

1. Choose a set of moves $\mathbf{s} \to \mathbf{s}'$ that visit all states *uniformly* and *reversibly*. That is, so that each microstate can eventually be reached from any other, and so that $P(\mathbf{s} \to \mathbf{s}') = P(\mathbf{s}' \to \mathbf{s})$. Flipping a randomly chosen spin is one example; another is the subject of Exercise 1B.

2. Now, start your simulation off in some configuration $\mathbf{s}$, and generate a **trial move** to a new microstate from the set defined in step 1.

3. If the transition to the new microstate causes the heat to be transferred to the heat bath (i.e., a loss of energy), accept it: that is, set $\mathbf{s}$ equal to this new microstate. Otherwise, accept the move only with probability $p = \mathrm{e}^{\Delta Q/kT}$, where $\Delta Q < 0$ is (minus) the energy received by the system from the bath. In practice this is achieved if a uniformly-distribution random number between 0 and 1 is less than p.

4. Go back to step 2 (but starting from the newly-accepted configuration).

This algorithm is called the **Metropolis algorithm** and is an example of a **Monte Carlo** sampling scheme. These schemes are used extensively in statistical mechanics: see Newman and Barkema (1999) for an excellent textbook on the subject.

We see directly how free energy minimisation is achieved through these stochastic dynamics. Without the rejection of energy increasing moves, all configurations are sampled with equal probability: entropy is maximised. By rejecting some fraction of these moves, the dynamics are biased towards moves that minimise the energy. Equilibrium is reached when these two competing factors balance out.

1.8 Identifying the heat transferred in a transition

The key step in implementing this procedure is identifying the heat transferred in a transition from microstate $\mathbf{s}$ to $\mathbf{s}'$. From the first law we have

$$\Delta Q = -(\Delta U + \Delta W) \tag{1.14}$$

where ΔQ is the heat absorbed by the bath, ΔW is the work done by the system against an applied field, and ΔU its internal energy change.

If we were simulating a fluid system of fixed volume, no work would be done in a transition, and $\Delta Q = -\Delta U$. On the other hand, if one has a magnetic system in a field, then when a spin is flipped then work is either done by or against the field. In this case, we have $\Delta W = -h\Delta M$, where h is the external field, and hence

$$\frac{P(\mathbf{s} \to \mathbf{s}')}{P(\mathbf{s}' \to \mathbf{s})} = \exp\left(-\frac{\Delta U - h\Delta M}{kT}\right) \ . \tag{1.15}$$

1.9 The equilibrium distribution of microstates

The detailed balance relation (1.13) is obtained by using the fact that the dynamics within the combined system are time-reversal symmetric, as will be seen in Section 1.11 below. For the system of interest, this implies that the *probability* of seeing a trajectory $\mathbf{s}_0 \to \mathbf{s}_1 \to \mathbf{s}_2 \to \cdots \to \mathbf{s}_n$ is equal to that of seeing the reverse trajectory $\mathbf{s}_n \to \mathbf{s}_{n-1} \to \cdots \to \mathbf{s}_0$. That is

$$\begin{aligned}
P(\mathbf{s}_0 \to \mathbf{s}_1 \to \mathbf{s}_2 \to \cdots \to \mathbf{s}_n) &= P(\mathbf{s}_n \to \mathbf{s}_{n-1} \to \cdots \to \mathbf{s}_0) && (1.16)\\
\implies P(\mathbf{s}_0)P(\mathbf{s}_0 \to \mathbf{s}_1)\cdots P(\mathbf{s}_{n-1} \to \mathbf{s}_n) &= P(\mathbf{s}_n)P(\mathbf{s}_n \to \mathbf{s}_{n-1})\cdots P(\mathbf{s}_1 \to \mathbf{s}_0) && (1.17)\\
\implies \frac{P(\mathbf{s}_n)}{P(\mathbf{s}_0)} &= \frac{P(\mathbf{s}_0 \to \mathbf{s}_1)\cdots P(\mathbf{s}_{n-1} \to \mathbf{s}_n)}{P(\mathbf{s}_1 \to \mathbf{s}_0)\cdots P(\mathbf{s}_n \to \mathbf{s}_{n-1})} \ , && (1.18)
\end{aligned}$$

where $P(\mathbf{s})$ is the equilibrium probability of configuration $\mathbf{s}$.

Using now the detailed balance relation (1.13), we find

$$\frac{P(\mathbf{s}_n)}{P(\mathbf{s}_0)} = \exp\left(\frac{\Delta Q}{kT}\right) \tag{1.19}$$

where now ΔQ is the total heat surrendered to the bath during the sequence of transitions $\mathbf{s}_0 \to \mathbf{s}_1 \to \cdots \mathbf{s}_n$. We thus find that the equilibrium distribution is the Boltmann distribution: that is, for a spin system in an external magnetic field we have that

$$P(\mathbf{s}) = \exp\left(-\frac{U(\mathbf{s}) - hM(\mathbf{s})}{kT}\right) . \tag{1.20}$$

Note that to obtain this result we have assumed only that all configurations in the combined system are equally likely, and that the dynamics are time-reversal symmetric. We can take these as the two defining properties of an equilibrium state.

By performing the Monte Carlo algorithm described in Section 1.7, we thus find that the number of times we visit a configuration $\mathbf{s}$ is proportional to the equilibrium probability $P(\mathbf{s})$. Put another way, configurations are sampled in the computer simulation with the same relative frequency that they would be in the corresponding physical system—albeit on a much slower overall timescale. Actually, to prove this result more formally, we would need to show that all configurations can eventually be reached in the simulation. We will not pursue this proof here, but we will see a simple example demonstrating convergence to the equilibrium distribution in Lecture 2.

1.10 The equilibrium distribution of macrostates

As we generate microstates using the Monte Carlo method, we can compute any statistical quantity of interest, such as the mean energy or magnetisation, or their fluctuations. In particular, we could count the number of times we visit a particular macrostate, for example, spin configurations with a given magnetisation $\tilde{M}$ (which need not equal the equilibrium magnetisation M). The resulting histogram would eventually converge onto the equilibrium **macrostate distribution function** $P(\tilde{M})$.

What use is such a function? If $\Omega(\tilde{M})$ is the number of microstates corresponding to the macrostate, then we will find that

$$P(\tilde{M}) \propto \Omega(\tilde{M}) \exp\left(-\frac{\tilde{U}(\tilde{M}) - h\tilde{M}}{kT}\right) = \exp\left(-\frac{\tilde{U}(\tilde{M}) - T\tilde{S}(\tilde{M}) - h\tilde{M}}{kT}\right) \tag{1.21}$$

where we have used the fact that $S = k \ln \Omega$ and assumed that the internal energy is a well-defined function of the magnetisation. Hence, we can measure the Gibbs free energy (1.9) in our simulation using

$$\tilde{G} = -kT \ln P(\tilde{M}) + \text{const} . \tag{1.22}$$

The behaviour of the macrostate distribution function is particularly interesting near the critical point. For the two-dimensional Ising model, it is known (see, e.g., Bruce and Wilding (1992)) that it exhibits the **scaling form**

$$P(\tilde{M})\big|_{T=T_c;h=0} \approx aN^{-15/16} P_s(a\tilde{M}N^{-15/16}) , \tag{1.23}$$

where a is some constant and N is the system size (number of spins). A consequence of this scaling form is that the fluctuations about the mean of the order parameter are of the same order in N as the order parameter itself (a fact that we will need to confront to better understand critical phenomena). The scaling form also implies that if one measures the macrostate distribution function, for example, by a Monte Carlo algorithm, for different system sizes and plots each distribution obtained as a function of the **scaling variable** $u = a\tilde{M}N^{-15/16}$, the different curves should lie on top of each other. Such data are shown in Figure 1.4, taken from Bruce and Wilding (1992).

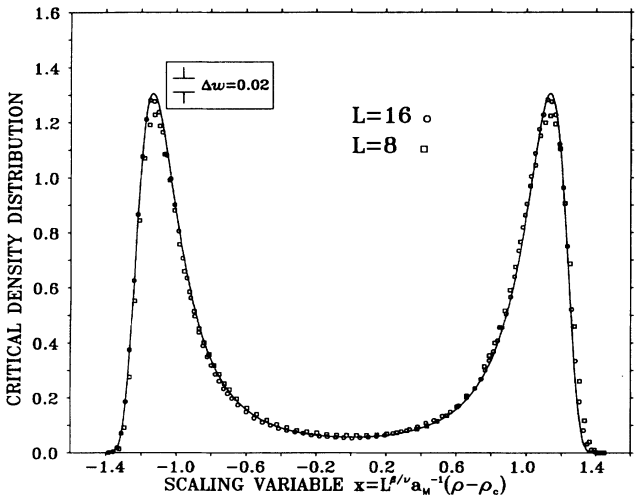


Figure 1.4 *Order parameter distribution at criticality for the Ising model (solid line) and Lennard-Jones fluid (points) in two dimensions. The data were obtained by Monte Carlo simulation, with the latter performed at two different linear system sizes $L = 8$ and 16 in units of 2σ. This corresponds to $N = L^2$ spins in the Ising model. Figure from Bruce and Wilding (1992).*

On the figure, the corresponding order parameter distribution for the Lennard-Jones fluid is plotted. The data not only fit the same scaling form as (1.23), where N is the number of particles and $\tilde{M}$ the density difference, with the same exponent $\frac{15}{16}$ but also the functional form of the scaling function $P_s(u)$ is the same for both the fluid and the magnet. (The only thing that is different is the parameter a). This provides very strong evidence for the universality of the ordering transition in the fluid and magnetic system. The origin of this universality will be revealed in Lecture 3.

1.11* Derivation of the detailed balance principle

We now derive the detailed balance relation (1.13) by invoking just two basic physical principles: (i) that at equilibrium, all microscopic configurations of the closed, combined system (system of interest + bath) are equally likely; (ii) the underlying dynamics (e.g., Hamiltonian dynamics) are time-reversal symmetric. The derivation is not difficult, although we need to set up some notation to complete it.

Consider a particular trajectory of the particles in the system of interest, lasting a time t, and call this X. This comprises the time evolution of all degrees of freedom within the system of interest—and we shall for the purposes of this derivation take the system to be a fluid system where the particles have positions and momenta. At the initial time, we can identify a set of possible bath configurations that lead to this trajectory being followed. Let us denote the size of this set (technically, the phase space volume) as $\Omega[X]$. See the shaded area in the upper left-hand part of Figure 1.5.

As the system evolves, this phase space volume is conserved (as the underlying dynamics are Hamiltonian), as indicated by the transition from the left to the right part of Figure 1.5. Let us now reverse all momenta (indicated by the vertical arrow in the figure). As the Hamiltonian is time-reversal symmetric (it depends only on even powers of momenta), subsequent evolution of the system over time t will recover the initial condition. Therefore, the number of configurations in the bath associated with the time-reversed trajectory $\hat{X}$ is the same as that associated with the forward trajectory:

$$\Omega[X] = \Omega[\hat{X}] \; . \tag{1.24}$$

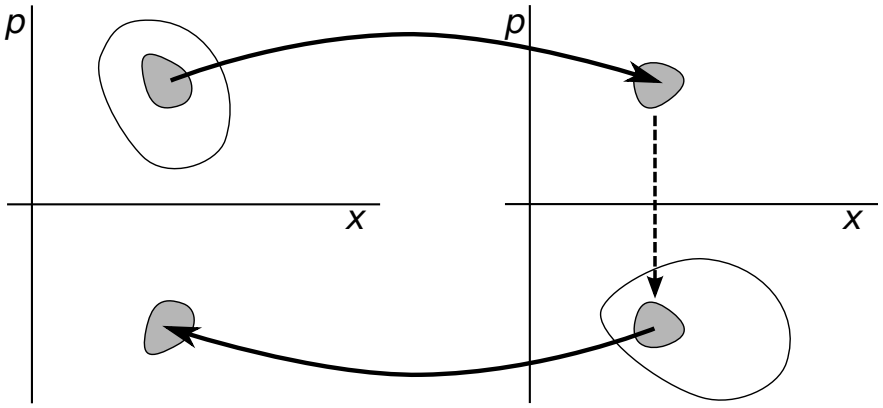


Figure 1.5 *Origin of detailed balance: the volume of phase space in the bath (shaded) along a trajectory in the system of interest is a conserved quantity. Larger areas correspond to phase space volumes associated with the endpoints of the trajectory.*

At equilibrium, all configurations of the combined system are equally likely. Therefore, the probability of witnessing the trajectory X, $P(X)$ is given by

$$P(X) = \frac{\Omega[X]}{\sum_{X'} \Omega[X']} = \frac{\Omega[\hat{X}]}{\sum_{\hat{X}'} \Omega[\hat{X}']} = P(\hat{X}) \ . \tag{1.25}$$

That is, the forward and reverse trajectories are equally likely to be observed at equilibrium.

Let us assume we know the starting configuration of the system of interest,X_0. What then is the probability that the specific trajectory X is followed (as opposed to any of the others that could be taken)? If $\Omega[X_0]$ is the total size of the set of *all* trajectories of the combined system where the system of interest starts at X_0, we have

$$P(X|X_0) = \frac{\Omega[X]}{\Omega[X_0]} = \frac{\Omega[\hat{X}]}{\Omega[X_0]} = \frac{\Omega[\hat{X}]}{\Omega[\hat{X}_0]} \frac{\Omega[\hat{X}_0]}{\Omega[X_0]} = P(\hat{X}|\hat{X}_0) \frac{\Omega[\hat{X}_0]}{\Omega[X_0]} \ . \tag{1.26}$$

Here, $\Omega[\hat{X}_0]$ is the number of bath configurations associated with the start of the time-reversed trajectory. If the combined system has zero net momentum, it follows that $\Omega[\hat{X}_0] = \Omega[X_t]$, the number of bath configurations associated with the end of the forward trajectory. Using Boltzmann's formula relating phase space volume to entropy, $S = k \ln \Omega$, and the Clausius relation $\Delta Q = T\Delta S$, we find the important relation between conditional probabilities:

$$\boxed{\frac{P(X|X_0)}{P(\hat{X}|\hat{X}_0)} = \exp\left(\frac{\Delta Q}{kT}\right)} \ . \tag{1.27}$$

This is the most general form of the detailed balance for the dynamics of an equilibrium system. Note that here ΔQ is the heat transferred *to* the heat bath over the course of the trajectory X.

Note that this relation applies to both fluid and spin systems. Spin systems are slightly simpler in that the a configuration and its time-reversed counterpart are identical. If we now take the trajectory $X = \mathbf{s} \to \mathbf{s}'$, the its time reversal is $\hat{X} = \mathbf{s}' \to \mathbf{s}$ and we recover (1.13). One implicit assumption we have made is that for longer trajectories, $P(\mathbf{s}_0 \to \mathbf{s}_1 \to \mathbf{s}_2)$ can be written as $P(\mathbf{s}_0 \to \mathbf{s}_1)P(\mathbf{s}_1 \to \mathbf{s}_2)$. That is, at each step, the probability of making a transition depends only on the current state of the system, and not any earlier history. Such a stochastic process is known as a **Markov process**.

Exercises

1A: Equivalence of Helmholtz and Gibbs free energies

The simplest magnetic system comprises N non-interacting spins $s_i = \pm 1$ with a potential energy $U(\mathbf{s}) = 0$ for all configurations. We will verify that the Helmholtz and Gibbs free energies provide equivalent descriptions of the thermodynamics of the system.

(a) If the total magnetisation $M = \sum_i s_i$ is constrained, show that in the thermodynamic limit ($N \to \infty$) the Helmholtz free energy per spin is

$$f(T, m) = \lim_{N\to\infty} \frac{F(T, M)}{N} = kT \left(\frac{1+m}{2} \ln \frac{1+m}{2} + \frac{1-m}{2} \ln \frac{1-m}{2} \right) ,$$

where m is the magnetisation per spin, $m = M/N$. [Hint: you will need to count the number of ways that a state with magnetisation M can be realised, write down the corresponding entropy and use Stirling's approximation $\ln n! \approx n \ln n - n$].

(b) Obtain the Gibbs free energy per spin $g(T, h)$, valid for the ensemble in which the magnetisation can fluctuate, by following through the sequence of steps described at the end of section 1.3, i.e., by obtaining the equation of state $h = \frac{\partial f(T,m)}{\partial m}$, inverting and substituting into $g(T, h) = f(T, m[T, h]) - hm(T, h)$. Show that this is formally equivalent to writing down the function $\tilde{g}(\tilde{m}, T, h) = f(T, \tilde{m}) - h\tilde{m}$ and minimising with respect to the unconstrained parameter $\tilde{m}$.

(c) Verify that the equation of state obtained from the Gibbs free energy via $m = -\frac{\partial g(T,h)}{\partial h}$ is the same as that found from the Helmholtz free energy.

1B: Discrete dynamics and detailed balance

(a) Consider a system of M particles on a closed $L \times L$ square lattice (i.e., particles cannot enter or leave at the boundaries). How many configurations can the system adopt if multiple particles may occupy a site? Show that the following algorithm generates a uniform distribution over these configurations and is reversible:

1. Choose a site with a probability proportional to its number of neighbours and the number of particles sitting on it.
2. Move a randomly-chosen particle on that site to a randomly-chosen neighbour.

Discuss the consequences of leaving out either (or both) of the factors in Step 1.

(b) Introduce now an attractive interaction between particles on the same site, so that the internal energy is $U(\mathbf{x}) = -\sum_i u(x_i)$ where x_i is the number of particles on site i. Verify that the Metropolis algorithm generates the desired equilibrium distribution $P(\mathbf{x}) \propto \mathrm{e}^{\sum_i u(x_i)/kT}$.

An alternative to the Metropolis algorithm is the **heat-bath algorithm**, where a particle move from $\mathbf{x} \to \mathbf{x}'$ is accepted with probability

$$\frac{1}{2} \left(1 - \tanh \frac{U(\mathbf{x}') - U(\mathbf{x})}{2kT} \right) .$$

Verify that this also satisfies detailed balance, and discuss the relative efficiency of the two algorithms.

1C: Wolff cluster algorithm for the Ising model

Local spin-flip dynamics can be used to simulate the Ising model, but can be inefficient, especially near a critical point, where many single spin flips are needed to realise the large-scale fluctuations seen near criticality. One way to speed up the simulations is to flip clusters of aligned spins. Show that the obvious choice, flipping entire domains of spins, can only simulate the Ising model at $T = 0$.

A cluster algorithm that works at nonzero temperature is ascribed to Wolff (1989).

1. Choose a site i at random, and mark this spin alone as belonging to a 'cluster', and designate all bonds between neighbouring spins as 'unmarked'.
2. For each unmarked bond connected to the cluster, mark it as 'active' with probability $p = 1 - \mathrm{e}^{-2J/kT}$ if the two spins it connects are aligned; otherwise, mark the bond 'inactive'.
3. Add to the cluster any sites connected to it by active bonds.
4. Go back to step 2, stopping once all bonds connected to the cluster are marked.

Discuss the $T = 0, T = \infty$ limits of this algorithm and show that it satisfies detailed balance with respect to the equilibrium distribution of the Ising model at finite T.

Lecture 2

A first attempt at modelling spontaneous order

In the first lecture we discussed the **spontaneous ordering**—freezing or alignment of spins—that occurs when the temperature of a system is reduced below some critical value. We introduced the **order parameter** ϕ that is zero when the energy of thermal fluctuations is sufficient to kick the system of interest out of energy minima, and nonzero when it isn't. We begin our quest for a deeper understanding of this phenomenon by looking at models that can be couched purely in terms of the global order parameter.

2.1 Spin dynamics in an isolated system

Let us return to the picture outlined in Lecture 1 where we had a combined system that is isolated from its environment. We will take this system to comprise N spins that can take values $s_i = \pm 1$, and have both the internal energy $U(\mathbf{s})$ and the magnetisation $M = \sum_i s_i$ to be conserved quantities. This can be achieved if every spin interacts with every other in an Ising-like manner:

$$U(\mathbf{s}) = -J \sum_{\langle i,j \rangle} s_i s_j = -\frac{J}{2} \left[\sum_{i=1}^{N} s_i \sum_{j=1}^{N} s_j - \sum_{i=1}^{N} s_i^2 \right] = -\frac{J}{2} \left[M^2 - N \right] \ . \tag{2.1}$$

Such an interaction is often described as **infinite range**.

Let us also define a dynamics for this isolated system where, in each time step, a random pair of interacting spins exchange spin states. (These dynamics are known as **Kawasaki dynamics**). Since all spins interact with each other, this corresponds to exchanging a random pair of spins on each tick of a clock. See Figure 2.1.

2.2 Dynamics of a single spin in contact with a bath

Let us now single out one of the spins (say, $i = 1$) and take it to be our system of interest that is in contact with a bath comprising the remaining $N - 1$ spins. Under the dynamics described above, this spin will repeatedly flip by interacting with the thermal environment provided by the bath of spins. We define the order parameter for the system of interest as its *average* (mean) magnetisation, i.e., $\phi \equiv m = \langle s_1 \rangle$. (We will in fact use the symbol m to denote the order parameter from now on).

To see how this spin behaves, we need to write down the probability that the spin flips from up $s_1 = 1$ to down $s_1 = -1$, or vice versa, in a single time step. Specifically, we seek first the probability that an up-spin flips *given* that it is initially pointing up. Since two spins are exchanged in each timestep, the

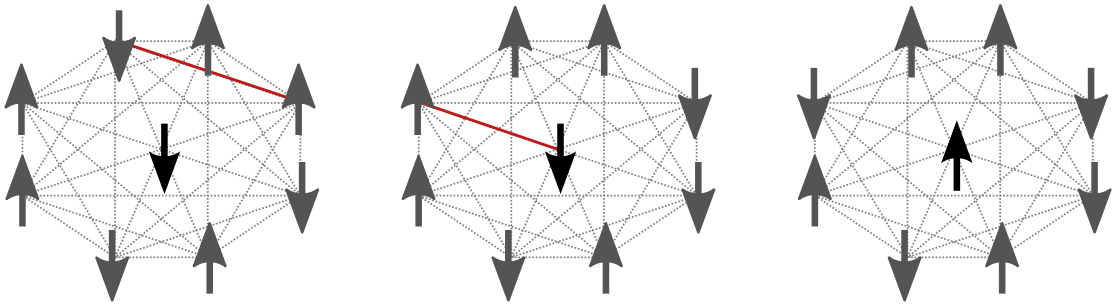


Figure 2.1 *Dynamics of an infinite-range spin model. In each timestep, a random pair of spins is chosen and their states exchanged. We are interested in the statistics of the central "spin of interest".*

probability that the chosen pair includes spin 1 is $\frac{2}{N}$. If it is pointing up, it flips with a probability equal to the fraction of down-spins in the bath. Now, if the total magnetisation is $M = n_\uparrow - n_\downarrow$ where $n_\uparrow$ and $n_\downarrow$ are the total number of up- and down-spins, $n_\uparrow + n_\downarrow = N$, we have that $n_\downarrow = \frac{N-M}{2}$, and hence that the fraction of down-spins in the bath is $\frac{N-M}{2(N-1)}$. So, the probability that an up-spin flips to down in one timestep is:

$$P(\uparrow\to\downarrow) = \frac{2}{N}\frac{N-M}{2(N-1)} = \frac{1-\frac{M}{N}}{N-1} \ . \tag{2.2}$$

By similar reasoning,

$$P(\downarrow\to\uparrow) = \frac{2}{N}\frac{N+M}{2(N-1)} = \frac{1+\frac{M}{N}}{N-1} \ . \tag{2.3}$$

2.3 Master equation for the spin relaxation

Let us say we know the spin's state at some time t. We can work out the probability that it has a particular state at time $t+1$ by writing down what is known as the **master equation** for the stochastic process. For a general Markov process with transition probabilities given by $P(\mathbf{s}\to\mathbf{s})$, as discussed in Lecture 1, the master equation (note lower-case 'm') reads

$$P(\mathbf{s}, t+1) = \sum_{\mathbf{s}'} P(\mathbf{s}', t) P(\mathbf{s}'\to\mathbf{s}) \ . \tag{2.4}$$

That is, the probability that we're in state $\mathbf{s}$ after one timestep is equal to the probability of being in some other state $\mathbf{s}'$ in the previous timestep, multiplied by the probability of making the transition to $\mathbf{s}$, and then summed over all possible previous states.

Note that the master equation is, in fact, a set of equations, one for each state. For the single-spin system that has two states, we therefore have two equations:

$$\begin{aligned} P(\uparrow, t+1) &= P(\downarrow, t)P(\downarrow\to\uparrow) + P(\uparrow, t)P(\uparrow\to\uparrow) & (2.5)\\ P(\downarrow, t+1) &= P(\uparrow, t)P(\uparrow\to\downarrow) + P(\downarrow, t)P(\downarrow\to\downarrow) \ . & (2.6) \end{aligned}$$

By $P(\uparrow\to\uparrow)$ and $P(\downarrow\to\downarrow)$ we mean the probability of staying in the up or down state in the course of one timestep. By conservation of probability, that is, $P(\uparrow\to\downarrow) + P(\uparrow\to\uparrow) = 1$, we have

$$\begin{aligned} P(\uparrow\to\uparrow) &= 1 - P(\uparrow\to\downarrow) = \frac{N-2+\frac{M}{N}}{N-1} & (2.7)\\ P(\downarrow\to\downarrow) &= 1 - P(\downarrow\to\uparrow) = \frac{N-2-\frac{M}{N}}{N-1} \ . & (2.8) \end{aligned}$$

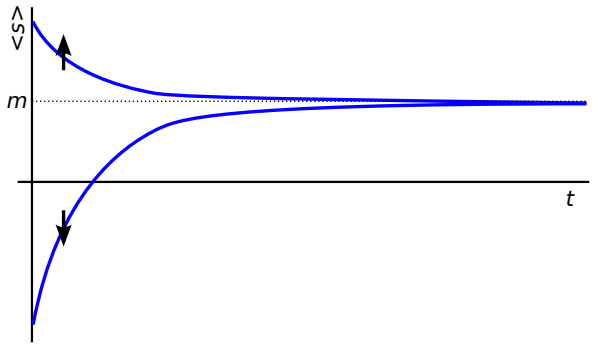


Figure 2.2 *The solution of the master equation for the spin of interest implies an exponential relaxation of the magnetisation of a site from its initial value $s(0) = \pm 1$ to the equilibrium value m.*

The master equation is most conveniently written as a matrix equation

$$\begin{pmatrix} P(\uparrow, t+1) \\ P(\downarrow, t+1) \end{pmatrix} = \begin{pmatrix} \frac{N-2+\frac{M}{N}}{N-1} & \frac{1+\frac{M}{N}}{N-1} \\ \frac{1-\frac{M}{N}}{N-1} & \frac{N-2-\frac{M}{N}}{N-1} \end{pmatrix} \begin{pmatrix} P(\uparrow, t) \\ P(\downarrow, t) \end{pmatrix} \tag{2.9}$$

or, more succinctly,

$$\mathbf{P}(t+1) = \mathsf{M}\mathbf{P}(t) \; . \tag{2.10}$$

Whenever we are faced with a matrix equation like this, our response is always the same: we diagonalise M. The eigenvalues are found in the usual way to be:

$$\lambda_1 = 1 \quad \text{and} \quad \lambda_2 = 1 - \frac{2}{N-1} \; . \tag{2.11}$$

The eigenvectors have a slightly complicated structure due to the fact that the matrix M is not symmetric—this will be explored in Exercise 2B. For now, we note that if $\mathbf{v}_i$ satisfies the equation $\mathsf{M}\mathbf{v}_i = \lambda_i \mathbf{v}_i$, and $P(0) = \mathbf{v}_i$, we have after time t that

$$P(t) = \mathsf{M}^t \mathbf{P}(0) = \mathsf{M}^t \mathbf{v}_i = \lambda_i^t \mathbf{v}_i \; . \tag{2.12}$$

Hence, the eigenvector $\mathbf{v}_1$ is preserved by the dynamics, and corresponds to the **stationary** (equilibrium) distribution. The eigenvector $\mathbf{v}_2$ decays exponentially with characteristic time

$$\theta_2 = -\frac{1}{\ln \lambda_2} - \frac{1}{\ln\left(1 - \frac{2}{N-1}\right)} \approx \frac{N}{2} \; . \tag{2.13}$$

This can be used to work out how a spin relaxes to its equilibrium distribution, as will be seen in Exercise 2B. Schematically the relaxation can be sketched as in Figure 2.2.

The increase of the relaxation time with N is due to the fact that as the size of the bath is increased, the probability that the spin of interest is chosen to flip decreases correspondingly. If we were simulating the single spin in contact with a large heat bath (see Lecture 1), the transition probabilities (2.2) and (2.3) would be a bad choice because most of the time, our simulation would do nothing. However, we could easily speed up the simulation by not bothering with the updates where both spins are in the bath, and performing only those where the spin in the system of interest changes sign. This amounts to rescaling both (2.2) and (2.3) by a constant factor that decreases the relaxation time, but leads to the same steady state overall—which is what you often want in a Monte Carlo simulation.

2.4 Equilibrium condition and phase transition

The stationary distribution for the system of interest is given by the vector $\mathbf{v}_1$. This is straightforwardly found (from the eigenvalue equation) to be

$$\begin{pmatrix} P_{\rm eq}(\uparrow) \\ P_{\rm eq}(\downarrow) \end{pmatrix} = \begin{pmatrix} \frac{1+\frac{M}{N}}{2} \\ \frac{1-\frac{M}{N}}{2} \end{pmatrix} . \tag{2.14}$$

The equilibrium magnetisation, or order parameter, is therefore

$$m = \langle s_1 \rangle = P_{\rm eq}(\uparrow) \times 1 + P_{\rm eq}(\downarrow) \times (-1) = \frac{M}{N} , \tag{2.15}$$

exactly as we expect, as this is the imposed magnetisation of the combined, isolated system.

How does this tell us anything about the spontaneous ordering that occurs at low temperatures? Recall that in Lecture 1, we derived the detailed balance relation

$$\frac{P(\mathbf{s} \to \mathbf{s}')}{P(\mathbf{s} \to \mathbf{s}')} = \exp\left[\frac{\Delta Q}{kT}\right] \tag{2.16}$$

where ΔQ is the heat absorbed by the bath in the transition $\mathbf{s} \to \mathbf{s}'$. In the absence of an external field, no work is done on the bath when a spin flip in the bath compensates that of the system of interest. Hence, ΔQ will equal the change to that part of the internal energy that comprises *only* the degrees of freedom in the bath. This division of the internal energy (2.1) is

$$U(\mathbf{s}) = -J \sum_{\langle i,j \rangle} s_i s_j = \underbrace{-J s_1 \sum_{i=2}^{N} s_i}_{\text{System of interest}} \quad \underbrace{-J \sum_{\langle i,j \rangle \; i,j>1} s_i s_j}_{\text{Bath}} . \tag{2.17}$$

By analogy with Eq. (2.1), we can write the internal energy of the bath as

$$U_b = -\frac{J}{2}\left[M_b^2 - N_b\right] \tag{2.18}$$

where M_b and N_b is the magnetisation and number of spins *in the bath*.

If there are $n_\uparrow$ up-spins in the combined system, one of which is the spin of interest, then in the bath the magnetisation is $M_b = 2n_{\uparrow,b} - (N-1) = 2(n_\uparrow - 1) - (N-1) = (2n_\uparrow - N) - 1 = M - 1$. Likewise, if the spin of interest is pointing down, $M_b = M + 1$. Hence, in the transition $\uparrow \to \downarrow$, M_b goes from $M-1$ to $M+1$ and the heat transferred to the bath in this transition is

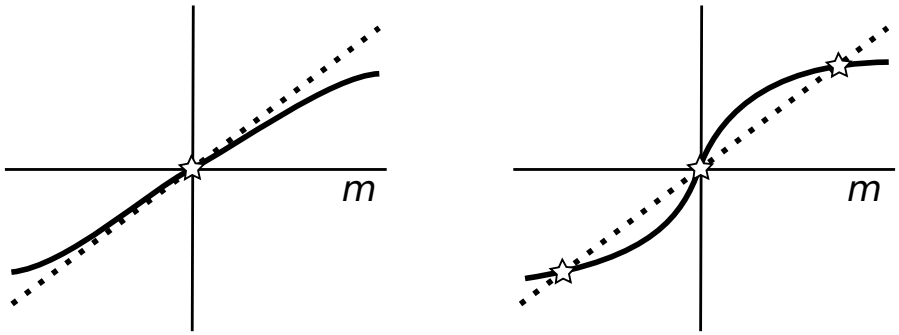
$$\Delta Q = \Delta U_b = -2JM . \tag{2.19}$$

Now, using (2.16), the transition probabilities (2.2) and (2.3), and this expression for the heat exchange, we obtain a criterion for the spin of interest to be in thermal equilibrium with the bath made up of the other spins:

$$\frac{1 - \frac{M}{N}}{1 + \frac{M}{N}} = \exp\left[-\frac{2JM}{kT}\right] . \tag{2.20}$$

We can read this equation in two ways: a fixed value of the magnetisation M implies a temperature T of the heat bath. Alternatively, we can view the temperature as an independent variable, and use this equation to fix the magnetisation self-consistently. We shall follow this latter path.

We remark that the condition (2.20) is *exact* for the model in which all spins interact with all others, because the heat transferred to the bath in the transition is independent of the state of the bath. In a model in which the spin(s) of interest interact only at the interface with the bath, this would no longer be true and (2.20) would hold only as an approximation.



(a) *High temperature* **(b)** *Low temperature*

Figure 2.3 *Solutions of the equation (2.21).*

2.5 The ordering transition

Recalling that our order parameter $m = \langle s_1 \rangle = \frac{M}{N}$, we get after substituting into (2.20) and rearranging that

$$m = \tanh\left[\frac{JNm}{kT}\right] \ . \tag{2.21}$$

From this, we can divine a **critical temperature** below which the system spontaneously orders. As this equation cannot be solved analytically for m, it is best to investigate it graphically, that is, by plotting the left- and right-hand sides on the same axes, and looking for points at which the curves cross—see Figure 2.3. At high temperatures, the line of gradient one intersects the tanh curve at $m = 0$. That is, half the spins are up, the other half down and this is the maximally disordered configuration of the system. At low temperatures, there are two other intersection points, $m = \pm m^*$. If we subject the global magnetisation (M) to external perturbations, we find that the disordered state with $m = 0$ is unstable to ordering of the spins, and hence these order states are the stable equilibrium states. See Exercise 2B.

The critical temperature can be found by asking the tanh curve to have a gradient of unity at the origin:

$$\frac{\mathrm{d}}{\mathrm{d}m}\tanh\left[\frac{JNm}{kT}\right]\bigg|_{m=0} = \frac{JN}{kT}\mathrm{sech}^2\left[\frac{JNm}{kT}\right]\bigg|_{m=0} = \frac{JN}{kT} = 1 \implies T_c = \frac{JN}{k} \ . \tag{2.22}$$

In Lecture 1, we stated that just below the critical point, the order parameter grows with the distance from the critical point as a power, $m \sim |T - T_c|^\beta$. To obtain a value for β, we introduce the reduced temperature

$$\tau = \frac{T - T_c}{T_c} \tag{2.23}$$

which is small near criticality. Since m is also small, the idea now is that we expand (2.21) in powers of m and τ to identify β:

$$m = \tanh\left[\frac{m}{1+\tau}\right] = \frac{m}{1+\tau}\left[1 - \frac{1}{3}\frac{m^2}{(1+\tau)^2} + \cdots\right] \tag{2.24}$$

or

$$0 = -m\tau - \frac{1}{3}m^3 + \cdots \ . \tag{2.25}$$

If $\tau > 0$, these first two terms can be balanced only if $m = 0$; when $\tau < 0$, the other solutions spring into existence, and one can also have $\tau \sim -\frac{1}{3}m^2$. That is, the exponent $\beta = \frac{1}{2}$.

There are other exponents that can be defined at the critical point, which we will come to shortly. These involve knowledge of other quantities, such as the response of the magnetisation to an applied

external field. In principle this could be calculated by following the procedure leading to (2.20), but by remembering to include the work done against the external field in (2.19)—see Eq. (1.14). This ultimately leads to the self-consistency relation

$$m = \tanh\left[\frac{JNm + h}{kT}\right] . \tag{2.26}$$

2.6 Landau theory for the free energy minimum and critical exponents

Rather than persevere with calculations for a specific model to introduce the remaining critical exponents, we will turn to a theory due to Landau which is more general and delivers the desired exponents more quickly to boot. The insight that delivers this extra power is that the critical exponents that arise depend on which powers of m and τ appear in expansions like (2.24). Since all physical quantities are derived from the free energy, the idea is to expand it as a power series around the critical point, discarding terms that are incompatible with the symmetries of the system, and keeping the minimum number of terms necessary to obtain sensible values for the exponents (as we did above).

We will continue with the case of an Ising ferromagnet, whose magnetisation m is determined by the temperature τ and the applied field h. As in Lecture 1, we will work with the Gibbs free energy, and initially treat the magnetisation as a free parameter $\tilde{m}$ whose equilibrium value will be fixed by minimising the free energy.

Within a mean-field approximation, we always have that

$$\tilde{g}(\tilde{m}, T, h) = u(\tilde{m}) - Ts(\tilde{m}) - h\tilde{m} \tag{2.27}$$

where the internal energy density $u(\tilde{m})$ and entropy density $s(\tilde{m})$ are both assumed to be fully determined by the global magnetisation $\tilde{m}$. (The reason why this is an approximation is expanded on below). Changing variable from T to τ, we have

$$\tilde{g}(\tilde{m}, \tau, h) = [u(\tilde{m}) - T_c s(\tilde{m})] - T_c \tau s(\tilde{m}) - h\tilde{m} . \tag{2.28}$$

The key feature of an Ising ferromagnet is that both the internal energy and the entropy are unchanged if we flip all the spins. This implies that $u(\tilde{m})$ and $s(\tilde{m})$ must both be invariant under $\tilde{m} \to -\tilde{m}$: that is, if we Taylor expand them about $\tilde{m} = 0$, we can only have even powers in the expansion.

Performing this expansion, we find

$$\tilde{g}(\tilde{m}, \tau, h) = \tilde{g}(0, 0, 0) + \frac{a(\tau)}{2}\tilde{m}^2 + \frac{b(\tau)}{4}\tilde{m}^4 - h\tilde{m} + O(\tilde{m}^6) \tag{2.29}$$

where $a(\tau)$ and $b(\tau)$ are (different) linear functions of τ, and $\tilde{g}(0, 0, 0)$ is a constant. Since the physics of the system does not depend on this constant, we can set it to zero. The factors of 2 and 4 have been included for convenience in the following. In principle, we could continue the expansion to higher order in $\tilde{m}$; however the terms we have retained here are sufficient to determine the critical exponents not just for the model discussed above, but all models that fall within the mean-field class.

It remains now to determine constraints on the linear functions $a(\tau)$ and $b(\tau)$. We are looking for only the $\tilde{m} = 0$ state to be stable (a free energy minimum) for $\tau > 0$, and for some state with some finite $\tilde{m} \neq 0$ to be stable for $\tau < 0$. In particular this means that we need $\tilde{g}(\tilde{m}, \tau, h)$ to increase as $\tilde{m} \to \pm\infty$, which means that we must have $b(\tau) > 0$ for all τ, and in particular at $\tau = 0$. Hence, for small τ, we will have $b(\tau) \approx b$ where b is a positive constant. Now, when $a(\tau)$ is positive, the free energy has

a single minimum (at $\tilde{m} = 0$), and when it is negative it has two minima at a nontrivial value of $\tilde{m}$. This tells us that the linear function $a(\tau)$ must change sign at $\tau = 0$, meaning that $a(\tau) \propto \tau$. We can take the constant of proportionality to equal 1 (since this is equivalent to choosing a temperature scale), which gives rise to the Landau series expansion

$$\tilde{g}(\tilde{m}, T, h) = \frac{\tau}{2}\tilde{m}^2 + \frac{b}{4}\tilde{m}^4 - h\tilde{m} \ . \tag{2.30}$$

If we now actually perform the minimisation to find the equilibrium magnetisation, we obtain

$$m(\tau, h) = \begin{cases} \left|\frac{\tau}{b}\right|^{1/2} & \tau < 0 \\ 0 & \tau > 0 \end{cases} \qquad \text{at } h = 0 \tag{2.31}$$

from which we recover the result $\beta = \frac{1}{2}$ obtained for the concrete model above. Note that to do so, we made only mild assumptions on how the free energy should vary in the vicinity of the critical point. Hence, the analysis here applies to all models for which these assumptions hold: different models will have a different constant b, which does not affect the critical exponent that is obtained.

It is now time to define the remaining critical exponents, all of which can be obtained by examining (2.29) in the appropriate limits (see Exercise 2A).

Specific heat	$c_h = T\frac{\partial s(T,h)}{\partial T} \sim \lvert\tau\rvert^{-\alpha}$	$h = 0$
Order parameter	$m \sim \lvert\tau\rvert^{\beta}$	$h = 0, \tau < 0$
Susceptibility	$\chi_T = \frac{\partial m(T,h)}{\partial h} \sim \lvert\tau\rvert^{-\gamma}$	$h = 0$
Critical isotherm	$h \sim \lvert m\rvert^{\delta}$	$\tau = 0$

(2.32)

No one can ever remember which one is which, except possibly β. Mnenomics on a postcard, please.

2.7 Scope of Landau free energy theory

Landau free energy theory is so beautifully simple, it would be lovely if it were exact. Unfortunately, as can be seen from Table 2.1, it can be seen that while experimental measurements of critical exponents in fluid systems agree rather well with those for the 3d Ising model (demonstrating again the universal nature of critical phenomena), they differ from the predictions from Landau theory.

The Landau free energy theory works for what is known as **mean field** models, an example of which we have examined extensively above. The essential property of such models is that their energy can be expressed purely in terms of the global order parameter (or, equivalently, the thermal average of

Exponent	Landau/Mean-field	2d Ising	3d Ising	Experiment
α	0	0	0.110	0.110–0.0116
β	$\frac{1}{2}$	$\frac{1}{8}$	0.325	0.316–0.327
γ	1	$\frac{7}{4}$	1.241	1.23–1.25
δ	3	15	4.82	4.6–4.9

Table 2.1 *Critical exponents for Ising-like systems, taken from Goldenfeld (1992). All numerical estimates for the 3d Ising model have an error in the last digit quoted.*

fluctuating quantities). For example, in the model where all spins interact with each other, the energy function could be written as being proportional to M^2, see Eq. (2.1) and compare with the energy term in the Landau expansion (2.29). More generally, mean-field theories neglect fluctuations and correlations in the order parameter, and it is this neglect that leads to the discrepancy between mean-field theory and reality. What Landau theory *does* achieve is to demonstrate that different models with the same basic symmetries (such as invariance under flipping all the spins) and other physical properties have the same exponents. This is the first step towards an explanation of universality.

A theory that takes fluctuations and correlations seriously is discussed in the next lecture. To conclude this lecture, we will establish what kind of free energy we are looking for, and further show that critical exponents are not independent quantities but are, in fact, related to each other.

2.8 Widom's scaling hypothesis

Recall that at a phase transition, the free energy exhibits discontinuities in its derivatives (at some, not necessarily, integer order). We focus on that part (the "singular" part) of the free energy that contains these discontinuities, and, following Widom, make the ansatz that it has the scaling form

$$g(\tau, h) = b^{-d} g(b^{y_1}\tau, b^{y_2}h) \tag{2.33}$$

where y_1 and y_2 are a pair of new exponents. Here b is interpreted as a spatial distance and d is the number of dimensions. The prefactor b^{-d} takes into account that if the unit of length is redefined by a change of scale b, the density will change accordingly (we will see more of this kind of thing in the next lecture). The key point is that because all thermodynamic properties are obtained by differentiating the free energy, all the critical exponents listed in the previous section must depend only on the scaling dimensions y_1 and y_2. Therefore, of these four exponents, only two of them are independent. Genuinely new exponents enter only when we consider quantities that are not derivable from the free energy as written here.

For example, let's consider the order parameter exponent. We have that $m \sim -\frac{\partial g(\tau,h)}{\partial h} \sim |\tau|^\beta$ for small negative τ and at $h = 0$. Using the ansatz (2.33) we have

$$m \sim -b^{-d+y_2} g_2(b^{y_1}\tau, 0) \ , \tag{2.34}$$

where by the subscript we mean "derivative w.r.t. the 2^{nd} argument". b is a free parameter, and is conveniently chosen such that $b^{y_1}\tau = 1$. Then

$$m \sim [-g_2(1,0)]\tau^{\frac{d-y_2}{y_1}} \ . \tag{2.35}$$

The stuff in square brackets is a constant we can hide under the $\sim$ sign. Hey presto, we have $\beta = \frac{d-y_2}{y_1}$. Following the same procedure for α, γ and δ and eliminating y_1 and y_2 we obtain exponent relations—over to you.

Exercises

2A: Critical exponents

(a) Using Widom's scaling ansatz (2.33), express the critical exponents α, β, γ and δ in terms of y_1 and y_2 and derive the relations $\alpha + 2\beta + \gamma = 2$ and $\beta(\delta - 1) = \gamma$. Verify that these relations hold for the data shown in Table 2.1.

(b) Minimise the Landau free energy (2.29) with respect to $\tilde{m}$ to obtain the equation of state

$$\tau m + bm^3 - h = 0 \ .$$

Use this to obtain the mean-field exponents β, γ and δ. Discuss the behaviour of the specific heat at criticality, and justify the conventional choice of the critical exponent $\alpha = 0$.

2B: Properties of the infinite-range spin model

(a) *Statics*—write down the internal energy and entropy of the isolated spin system, and therewith the Helmholtz free energy density

$$f(T,m) = \lim_{N\to\infty} \frac{U(M,N) - TS(M,N)}{N} \ .$$

What is the condition on J such that this free energy density is finite?

By transforming to the Gibbs free energy $g(T,h)$ (see Exercise 1A), obtain the equation of state valid for nonzero external field:

$$h = \frac{kT}{2} \ln\left(\frac{1+m}{1-m}\right) - JNm \ .$$

Sketch the isotherms in the m–h plane, and show that the susceptibility $\chi_T = \frac{\partial m}{\partial H} < 0$ at $h = m = 0$ below the critical temperature. Hence argue that the system is unstable at this point below T_c.

(b) *Dynamics*—in the lectures we derived the stochastic dynamics of a single spin interacting with a heat bath by random spin exchange. This led to a master equation which can be written in general terms as

$$\mathbf{P}(t+1) = \mathsf{M}\mathbf{P}(t)$$

where M is the 2×2 matrix

$$\mathsf{M} = \begin{pmatrix} 1-b & a \\ b & 1-a \end{pmatrix} \ .$$

Show that this matrix has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 1 - a - b$.

Find the **left eigenvectors** $\mathbf{u}_i$ and **right eigenvectors** $\mathbf{v}_i$ of this matrix, i.e., solutions of

$$\begin{aligned} \mathsf{M}\mathbf{v}_i &= \lambda_i \mathbf{v}_i \\ \mathbf{u}_i^{\mathrm{T}}\mathsf{M} &= \mathbf{u}_i^{\mathrm{T}}\lambda_i \end{aligned}$$

where the superscript T means transpose. Show that although the left- and right eigenvectors are not colinear, one can nevertheless normalise them such that the **biorthogonality** relation

$$\mathbf{u}_i \cdot \mathbf{v}_j = \delta_{i,j}$$

is satisfied. Hence, given initial condition $\mathbf{P}(0)$, show that

$$\mathbf{P}(t) = (\mathbf{u}_1 \cdot \mathbf{P}_0)\mathbf{v}_1 + \lambda_2^t (\mathbf{u}_2 \cdot \mathbf{P}_0)\mathbf{v}_2 \ .$$

Interpret the meaning of the eigenvector $\mathbf{v}_1$ and eigenvalue λ_2.

Using the values of a and b appropriate for the heat bath discussed in the lecture, $a = \frac{1+m}{N-1}$, $b = \frac{1-m}{N-1}$, where m is the mean magnetisation, show that the equilibrium spin autocorrelation function

$$\begin{aligned} \langle s(t)s(0)\rangle - \langle s(t)\rangle\langle s(0)\rangle &= \sum_{s,s'} s' P(s',t|s,0) s P_{\mathrm{eq}}(s) - \left(\sum_s s P_{\mathrm{eq}}(s)\right)^2 \\ &= (1-m^2)\left(1 - \frac{2}{N-1}\right)^t \ , \end{aligned}$$

and hence that in the limit $N \to \infty$,

$$\langle s(t)s(0)\rangle - \langle s(t)\rangle\langle s(0)\rangle =\sim (1-m^2)\mathrm{e}^{-2t/N} \ .$$

2C: Exactly solvable model with a phase transition

A class of two-dimensional interfaces can be modelled as a random walk that takes steps $(\delta x, \delta y) = (1, \pm 1)$. A model that exhibits a phase transition can be constructed for such an interface that starts at $(x,y) = (0,0)$, ends at $(x,y) = (2n,0)$ and is constrained such that for $0 < x < 2n, y \geq 0$. Sketch a realisation of the interface that contacts the origin m times (not including the contact at $(x,y) = (0,0)$). We will associate an energy $-\epsilon$ with each of these contacts and show that in the thermodynamic limit $n \to \infty$ there is a critical temperature T_c below which the interface is pinned to the origin (i.e., has a nonzero mean value of the order parameter $\phi = m/n$).

The free energy density of this system is $\tilde{f} = -\tilde{\phi}\epsilon - Ts(\tilde{\phi})$ where the entropy density is defined as

$$s(\tilde{\phi}) = k \lim_{n\to\infty} \frac{1}{n} \ln \Omega(n, n\tilde{\phi})$$

and the number of walks of length $2n$ contacting the origin m times is

$$\Omega(n,m) = \frac{m}{n} \binom{2n-m-1}{n-m} .$$

Minimise this free energy and show there is a continuous phase transition at $T_c = \frac{\epsilon}{k \ln 2}$ with a critical exponent $\beta = 1$.

You may be interested in calculating $\Omega(n,m)$. To do this, consider its generating function (discrete Laplace transform)

$$G(z,m) = \sum_{n=0}^{\infty} \Omega(n,m) z^n .$$

The way to read this generating function is that the power of z counts the length of walk. Show that $G(z,m_1)G(z,m_2) = G(z,m_1+m_2)$, and hence that $G(z,m) = G(z,1)^m$. Putting $G(z) = G(z,1)$ show, by considering the nesting of walks beginning and ending at height $y = 1$ within a single walk beginning and ending at height $y = 0$ that

$$G(z) = z\left[1 + G(z) + G(z)^2 + G(z)^3 + \cdots\right] = \frac{z}{1 - G(z)} .$$

To invert the generating function we make use of a trick called the **Lagrange inversion formula** which says that if $G(z)/z = f(G)$ (that is, the right-hand side contains z only implicitly through G) with $f(0) = 1$, the coefficient of z^n in the generating function $g(G)$ (where g is an arbitrary function of G) is given by the coefficient of G^{n-1} in the expression $\frac{1}{n} g'(G) f(G)^n$. Use this formula to calculate $\Omega(n,m)$. For a proof of the Lagrange inversion formula, see p182 of Wilf (2006).

Lecture 3

Equilibrium fluctuations and correlations

A more complete picture of phase transitions and critical phenomena can be obtained if we allow the order parameter to be spatially varying and exhibit correlations. The lengthscale characterising these correlations plays an important part in the theory.

3.1 What is a correlation?

If knowledge of one random variable gives you information about another, they are said to be **correlated**. (A manifestation of this idea in quantum mechanics is **entanglement**). In statistical mechanics, we are often interested in knowing how fluctuations in a random variable about its average at two different points are related. For spin systems, we consider **correlation functions** of the form

$$C_{ij} = \langle (s_i - \langle s_i \rangle)(s_j - \langle s_j \rangle) \rangle \ . \tag{3.1}$$

If spin j typically exceeds its mean value whenever spin i does, fluctuations about the mean magnetisation about these two sites are said to be positively correlated. On the other hand, spin j might prefer to point down when spin i is pointing up, in which case they would be negatively correlated.

Multiplying out the correlation function we find the alternative form

$$C_{ij} = \langle s_i s_j \rangle - \langle s_i \rangle \langle s_j \rangle \ . \tag{3.2}$$

If spin j is independent of spin i, the correlation function vanishes. Typically we expect this to be the case for spins separated by a very large distance, and for the correlation function to vanish with distance as

$$C(r) \sim \mathrm{e}^{-r/\xi} \tag{3.3}$$

where r is the distance between two spins. The characteristic decay length ξ is called the correlation length. See Figure 3.1 for an illustration of a case where the correlation length can be identified with a type of fluctuation.

It is possible to design an extension to Landau theory to obtain an expression for the correlation function for mean field-like models, see Exercise 3C. Furthermore, one can show that ξ diverges as the critical point as

$$\boxed{\xi \sim |\tau|^{-\nu}} \tag{3.4}$$

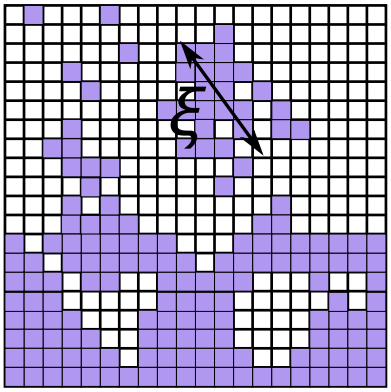


Figure 3.1 *Typical configuration of a 2d Ising model quenched to a subcritical temperature. An interface forms between domains of different spin states. Within these regions spontaneous fluctuations generate islands of a characteristic size that we can identify with the correlation length.*

where the exponent $\nu = \frac{1}{2}$ in mean-field theory. One can actually observe this divergence experimentally: you may have seen the phenomenon of **critical opalescence** demonstrated. As a binary fluid is taken through the critical temperature at which it demixes (phase separates). The mixture becomes cloudy as a consequence of lengthscale characterising density fluctuations having grown to the wavelength of visible light. It is this divergence we seek to understand in this lecture.

3.2 Manipulating the characteristic length scale: The renormalisation group

Our aim is to understand how systems with different correlation lengths—and in particular their free energies—are related. We therefore need to find a way to manipulate the correlation length.

This can be achieved by noting that there are two length-scales in a problem. One is given by the correlation length ξ, and the other is a microscopic lengthscale. For the case of the $2d$ Ising model, this would be the lattice spacing. Physically, only the ratio of these two lengthscales should be important, and so we can change this ratio by *coarse-graining* over microscopic degrees of freedom. Concretely, this can be achieved by grouping together spins into blocks of linear dimension b (so b^d spins per block, in a d-dimensional system) to form 'super-spins' (see Figure 3.2). This has the effect of increasing the lattice spacing by a factor of b whilst leaving the correlation length unchanged. Relative to the lattice spacing, however, the correlation length decreases by a factor b: $\xi' = \xi/b$. This is one of a number of changes that occur when this **renormalisation group** transformation is performed. Let us enumerate them.

Physical quantity	Value before transformation	Value after transformation
Number of spins	N	$N' = Nb^{-d}$
Correlation length	ξ	$\xi' = \xi/b$
Gibbs free energy	G	$G' = G$
Gibbs free energy density	$g = G/N$	$g' = G'/N' = (G/N)(N/N') = b^d g$
Temperature	T	T'
Magnetic field	h	h'

The last few of these relations need further explanation. When we initially perform the coarse-graining step, we are supposed to have two descriptions of the same system, so the total Gibbs free energy should be unaffected. However, because the number of spins has decreased, the free energy density has increased by the factor b^d. Moreover, equilibrium correlation length should depend on the thermodynamic

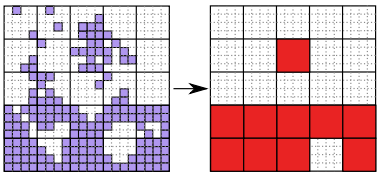


Figure 3.2 *Renormalisation of a 2d Ising model using a 'block-spin' transformation. Each* 4×4 *block of spins (filled squares, left) is replaced with a single up-spin (filled squares, right) if more than* 8 *spins in the original block are pointing up. Similarly, if more than* 8 *spins in the original block are pointing down (empty squares, left), they are replaced with a single down-spin (empty squares, right). In the case where the number of up- and down-spins are equal in a block, the choice of whether the super-spin points up or down is decided by a coin toss.*

constraints externally imposed on the system, that is, the values of T and h. Since the correlation length has changed, this implies T and h have also changed to new values T' and h'. In other words, the free energy density after the transformation $g'(T', h')$ should have the same functional dependence on T and h as the free energy before the transformation, that is, $g'(T', h') = g(T', h')$. But from the above table we have that $g'(T', h') = b^d g(T, h)$, or

$$g(T, h) = b^{-d} g'(T, h) = b^{-d} g(T', h') \ . \tag{3.5}$$

Notice that we would recover the Widom scaling form (2.33) if we were to find that $T' = b^{y_1} T$ and $h' = b^{y_2}$. By analysing the effect of the renormalisation group transformation on T and h, we will find that this is the case.

3.3 Fixed points and divergence of the correlation length

Formally, the renormalisation group defines a mapping $(T, h) \to (T', h')$ in a space of fields (or **coupling constants**) that depends on the rescaling parameter b. The **fixed points** of this transformation, that is, points where $(T', h') = (T, h)$, are of particular significance. In this section, we briefly explain why this is the case.

First of all, the existence of a fixed point might appear paradoxical at first sight. If T and h are unchanged, then ξ must also be unchanged, as ξ is a function of T and h. However, by design, we have asked for ξ to decrease by the factor b. The only way out is if at these fixed points we have $\xi = 0$ or $\xi = \infty$.

The former case implies no correlations between neighbouring spins, which is what one would expect at very low or very high temperatures. The latter case corresponds to the divergence of the correlation length at criticality that we have been looking for. Therefore, to understand critical phenomena, we need to understand fixed points of the renormalisation group transformation.

For example, if we are near to the critical point, we will find that the correlation function decreases under coarse-graining, as we would normally expect. Therefore, the critical point is a **repulsive** fixed point, and the high- and low-temperature fixed points are attractive. The rate of repulsion from the critical point is sufficient to determine the exponent ν. Recalling that $\xi \sim |\tau|^{-\nu}$ near the critical point, we have that

$$b = \frac{\xi}{\xi'} = \left| \frac{\tau}{\tau'} \right|^{-\nu} \ , \tag{3.6}$$

or, inverting,

$$\boxed{\nu = \frac{\ln b}{\ln \left| \frac{\tau'}{\tau} \right|}} \ . \tag{3.7}$$

Below we will show that the parameters y_1 and y_2 that appear in the Widom scaling expression (2.33) can be related to the rate of repulsion from a fixed point of the renormalisation transformation.

3.4* Example of an exactly renormalisable model

In setting up the renormalisation group transformation, we have assumed that the before and after the transformation, the system can be described by the same free energy function. More precisely, we begin with some set of microscopic degrees of freedom, for example, the spin configuration $\mathbf{s}$, and move to a coarse-grained description $\mathbf{s}'$, where a specific superspin state (say 'up') corresponds to multiple underlying spin configurations (say 'the number of up-spins in the block exceeds the number of downspins'). The probability distribution of configurations in the original model can be written as

$$P(\mathbf{s}) = \frac{1}{Z}\exp\left(-\frac{U(\mathbf{s}) - hM(\mathbf{s})}{kT}\right) . \tag{3.8}$$

The claim is that if we now replace $\mathbf{s}$ with the coarse-grained version of the configuration, $\mathbf{s}'$, the resulting probability distribution $P'(\mathbf{s})$ takes the same functional form, viz,

$$P'(\mathbf{s}') = \frac{1}{Z'}\exp\left(-\frac{U(\mathbf{s}') - h'M(\mathbf{s})}{kT'}\right) \tag{3.9}$$

where the temperature T and field h both change (they are 'renormalised') *but* the energy U and magnetisation M are fixed functions of the spin configuration.

This seems like a rather stringent requirement, and one that is unlikely to be true in general. Indeed, this is the case, and some care is needed to handle the fact that the functional form of the distribution does change under renormalisation. We will see how this is handled below. Nevertheless, there are models that are exactly renormalisable, one of which is the one-dimensional Ising model.

Instead of working with individual configurations, we will look at how the partition function Z, defined as

$$Z = \sum_{\mathbf{s}} \exp\left(-\frac{U(\mathbf{s}) - hM(\mathbf{s})}{kT}\right) , \tag{3.10}$$

changes under the renormalisation transformation. If the individual probabilities retain their functional form under the transformation, then so will the partition function. Consequently, we will be looking for Z and Z' to be expressible in terms of the same energy and magnetisation functions. For the 1d Ising model, these are $U(\mathbf{s}) = -J\sum_{j=1}^{N} s_j s_{j+1}$ and $M(\mathbf{s}) = \sum_{j=1}^{N} s_j$. Then,

$$Z = \sum_{\mathbf{s}} \exp\left(\frac{J\sum_{j=1}^{N} s_j s_{j+1} + h\sum_{j=1}^{N} s_j}{kT}\right) \tag{3.11}$$

and we will assume periodic boundary conditions $s_{i+N} = s_i$. It is helpful to introduce the function

$$w_{i,j}(K_0, K_1, K_2) = K_0 + K_1 s_i s_j + \frac{1}{2}K_2(s_i + s_j) \tag{3.12}$$

so that with $K_0 = 0$, $K_1 = J/kT$ and $K_2 = h/kT$ the partition function can be written as

$$Z = \sum_{\mathbf{s}} \exp\left[\sum_{j=1}^{N} w_{j,j+1}(K_0, K_1, K_2)\right] . \tag{3.13}$$

The parameters K_0, K_1 and K_2 are called **coupling constants**, and we want to see how they change under coarse-graining. (The reason for introducing K_0 will become clear shortly).

Our course-graining operation is to sum over even-numbered spins, leaving a sum over the odd-numbered spins. This corresponds to a rescaling by a factor $b = 2$. The aim is to arrive at a partition function in the form

$$Z' = \sum_{s_i,\, i \text{ odd}} \exp\left[\sum_{\text{odd } j} w_{j,j+2}(K_0', K_1', K_2')\right] \tag{3.14}$$

such that $Z = Z'$ for a suitable choice of (K_0', K_1', K_2'). To achieve this, we will require for any even-numbered spin i that

$$\sum_{s_i=\pm 1} \exp\left[w_{i-1,i}(K_0, K_1, K_2) + w_{i,i+1}(K_0, K_1, K_2)\right] = \exp\left[w_{i-1,i+1}(K_0', K_1', K_2')\right] \ . \tag{3.15}$$

Performing the sum, taking the log and using (3.12), this requirement becomes

$$2K_0 + \frac{1}{2}K_2(s_{i-1}+s_{i+1}) + \ln 2\cosh[K_1(s_{i-1}+s_{i+1})+K_2] = K_0' + K_1' s_{i-1}s_{i+1} + \frac{1}{2}K_2'(s_{i-1}+s_{i+1}) \ . \tag{3.16}$$

On the face of it, these appear to be two completely different functional forms. It turns out that there *is* a choice of (K_0', K_1', K_2') that renders them identical, and it can be discovered by asking for the two sides of the equation to agree at the physical values $s_{i-1}, s_{i+1} = \pm 1$. Enumerating the four possibilities yields three independent equations to solve simultaneously:

$$\begin{aligned}
K_0' + K_1' + K_2' &= 2K_0 + \ln 2 + \ln\cosh(2K_1 + K_2) + K_2 && (3.17)\\
K_0' - K_1' &= 2K_0 + \ln 2 + \ln\cosh K_2 && (3.18)\\
K_0' + K_1' - K_2' &= 2K_0 + \ln 2 + \ln\cosh(-2K_1 + K_2) - K_2 \ . && (3.19)
\end{aligned}$$

The fact that there are three equations explains why we needed three coupling constants in the Hamiltonian. Solving the linear system yields

$$\begin{aligned}
K_0' &= 2K_0 + \ln 2 + \frac{1}{4}\ln\left[\cosh(2K_1 + K_2)\cosh^2(K_2)\cosh(2K_1 - K_2)\right] && (3.20)\\
K_1' &= \frac{1}{4}\ln\left[\frac{\cosh(2K_1 + K_2)\cosh(2K_1 - K_2)}{\cosh^2(K_2)}\right] && (3.21)\\
K_2' &= K_2 + \frac{1}{2}\ln\left[\frac{\cosh(2K_1 + K_2)}{\cosh(2K_1 - K_2)}\right] && (3.22)
\end{aligned}$$

On their own these equations aren't particularly edifying, although we can make a few observations:

1. The 1d Ising model is *exactly* renormalisable within the space of coupling constants (K_0, K_1, K_2). By this we mean that the renormalisation transformation maps the 1d Ising model onto itself with modified values of a small number of coupling constants, and does not generate any new couplings (see the next section).

2. The parameter K_0 does not contain any physics—it just generates an additive constant in the free energy which vanishes when derivatives are taken to obtain physical quantities.

3. In zero field, $K_2 = 0$, we find $K_2' = 0$, that is, no coupling to the field is generated under the renormalisation transformation (as one might expect). The renormalisation of the spin-spin coupling then becomes

   $$K_1' = \frac{1}{2}\ln\cosh 2K_1 \ . \tag{3.23}$$

   As can be seen from Figure 3.3, iterating this equation leads to $K_1 = 0$, which corresponds to the infinite-temperature fixed point, for any starting value of $K_1 < \infty$. The fact that the critical

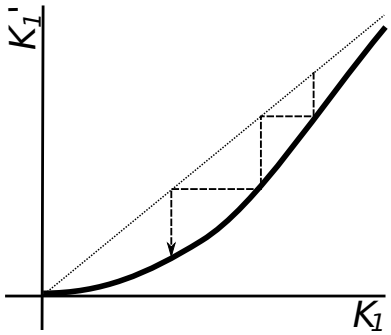


Figure 3.3 *Flow of the parameter $K_1 \to 0$ in the Migdal-Kandanoff renormalisation scheme applied to the 1d Ising model.*

fixed point lies at $K_1 = \infty$ ($T = 0$) makes the analysis a little more complicated. We shall use the quantity $X = \mathrm{e}^{-4K_1}$ as a surrogate for the temperature T near $T = 0$, since $X \to 0$ as $T \to 0$. Then, the renormalisation can be rewritten as

$$X' = \frac{4X}{(1+X)^2} \tag{3.24}$$

which has fixed points at $X = 0, 1$. The fixed point at $X = 0$ ($T = 0$) is clearly repulsive, and that at $X = 1$ ($T = \infty$) attractive. Near $X = 0$ we find that $X' \approx 4X$; putting $\tau \to X$ in (3.7) yields the critical exponent $\nu = \frac{1}{2}$.

3.5 Origin of universality

As we have seen, the renormalisation transformation rescales the correlation length and free energy, and generates a 'flow' of parameter T and h (or, more generally, coupling constants). If we are lucky, as in the 1d Ising model, we find that the transformation is closed in the space of coupling constants: it does not generate new coupling constants. In practice, this exact renormalisation is rare: each successive transformation can generate ever more couplings. A flavour of this was evident in the case of 1d Ising model described in the previous section: we needed to add the 'dummy' coupling constant K_0 alongside the physical ones K_1 and K_2 to render the model exactly renormalisable. In 2d the situation gets worse: applying the same scheme generates in the first instance next-nearest-neighbour interactions and couplings between quartets of spins that were not in the original energy function.

To understand how to handle this, let us formulate the generic problem we have to deal with. We begin by relating the free energy to the partition function:

$$\mathrm{e}^{-Ng(T,h)/kT} = Z = \sum_{\mathbf{s}} \mathrm{e}^{-H(\mathbf{s})} \tag{3.25}$$

where the quantity that appears in the Boltzmann weight is

$$H(\mathbf{s}) = K_1 A_1(\mathbf{s}) + K_2 A_2(\mathbf{s}) + K_3 A_3(\mathbf{s}) + \ldots \tag{3.26}$$

where the K_i are coupling constants and $A_i(\mathbf{s})$ are observables (also known as operators) that they couple to. For example, in the case of the Ising model, we have

$$H(\mathbf{s}) = \frac{J}{kT} \sum_{\langle i,j \rangle} s_i s_j + \frac{h}{kT} \sum_i s_i \tag{3.27}$$

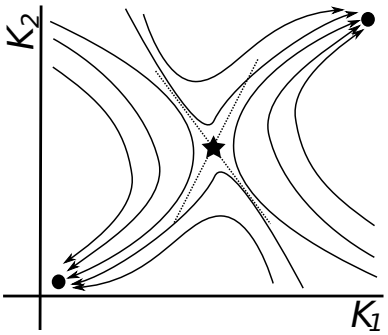


Figure 3.4 *Renormalisation group flow in a space spanned by two coupling constants. The repulsive fixed point ⋆ sends the flow towards one of the two attractive fixed points ●. Dashed straight lines show the eigendirections of the unstable fixed point.*

where $\langle i,j\rangle$ denotes nearest neighbour pairs. Hence

$$K_1 = \frac{J}{kT}\ , \quad A_i(\mathbf{s}) = \sum_{\langle i,j\rangle} s_i s_j \tag{3.28}$$

$$K_2 = \frac{h}{kT}\ , \quad A_i(\mathbf{s}) = \sum_{i} s_i \tag{3.29}$$

in this case. Different models would have different combinations of coupling constants and operators. The key point is that through (3.25) we can see that the free energy can be written some function of the coupling constants K_i:

$$g = g(K_1, K_2, K_3, \ldots)\ . \tag{3.30}$$

Under the renormalisation transformation, the K_i are mapped onto new values K_i'—for example, some that are originally zero may become nonzero. Our interest now is in the path through the space of coupling constants is generated by repeated application of the renormalisation group transformation.

The renormalisation transformation can be written as

$$K_i' = F_i(K_1, K_2, K_3, \ldots)\ , \tag{3.31}$$

and generates a flow that is illustrated schematically in Figure (3.4). Near a fixed point where $K_i' = K_i = K_i^*$, we can write $K_i = K_i^* + \delta K_i$. The δK_i then transform as

$$\delta K_i' = \sum_j \left.\frac{\partial F_i}{\partial K_j}\right|_{K^*} \delta K_j + O(\delta K^2)\ . \tag{3.32}$$

If we define the matrix M through

$$M_{ij} = \frac{\partial F_i}{\partial K_j} \tag{3.33}$$

we find that its eigenvalues and eigenvectors give insight into how the fixed points are entered or left.

If an eigenvalue of M has magnitude less than one, that describes an attraction to the fixed point: small displacements δK_i from K_i^* are then reduced in amplitude along the corresponding eigendirection. On the other hand, an eigenvalue larger than one describes a repulsion from the fixed point. A fixed point with mixture of attractive and repulsive eigendirections attracts a whole host of models all with initially different coupling constants, but sends them all away in the same directions as shown in Figure 3.4. It is this direction of repulsion that is common to a large class of models that underlies **universality**: critical exponents (as we have seen) are related to the rate of repulsion from a critical fixed point.

3.6 Derivation of the Widom scaling form

To recover the Widom scaling form for the free energy, we can write (3.26) as

$$H(\mathbf{s}) = H^*(\mathbf{s}) + \sum_i \delta K_i A_i(\mathbf{s}) \tag{3.34}$$

where $H^*(\mathbf{s}) = \sum_i K_i^* A_i(\mathbf{s})$. After the transformation,

$$H'(\mathbf{s}) = H^*(\mathbf{s}) + \sum_{i,j} A_i(\mathbf{s}) M_{ij} \delta K_j \ . \tag{3.35}$$

Decomposing onto the (normalised) eigenvectors $|\lambda\rangle$ of M, we can write the identity I and M as

$$\mathsf{I} = \sum_\lambda |\lambda\rangle\langle\lambda| \quad \text{and} \quad \mathsf{M} = \sum_\lambda |\lambda\rangle\lambda\langle\lambda| \tag{3.36}$$

so that (3.34) and (3.35) become

$$H(\mathbf{s}) \ = \ H^*(\mathbf{s}) + \sum_\lambda \langle \delta K|\lambda\rangle\langle\lambda|A(\mathbf{s})\rangle \tag{3.37}$$

$$H'(\mathbf{s}) \ = \ H^*(\mathbf{s}) + \sum_\lambda \langle \delta K|\lambda\rangle\lambda\langle\lambda|A(\mathbf{s})\rangle \ . \tag{3.38}$$

That is, in the vicinity of the fixed point, we can write

$$H(\mathbf{s}) = H^*(\mathbf{s}) + \sum_\lambda a_\lambda \langle\lambda|A(\mathbf{s})\rangle \tag{3.39}$$

where the a_λ parameters (which are linear combinations of the K_i) transform as

$$a'_\lambda = \lambda a_\lambda \ . \tag{3.40}$$

Collectively, these parameters specify the temperature, magnetic field and other coupling constants for the model. So we can write the free energy as a function of these a_λ parameters. Earlier we found the transformation rule for the free energy to be

$$g(a_{\lambda_1}, a_{\lambda_2}, \ldots) = b^{-d} g(a'_{\lambda_1}, a'_{\lambda_2}, \ldots) = b^{-d} g(\lambda_1 a_{\lambda_1}, \lambda_2 a_{\lambda_2}, \ldots) \tag{3.41}$$

which is one step closer to the Widom form. To get even closer, we observe that the magnitude of the eigenvalues depends on the chosen scale factor b. However, we must have that the renormalised Hamiltonian obtained by rescaling by a factor b n times in succession agrees with that obtained by rescaling by a factor b^n once. This imposes the constraint that

$$\lambda(b^n) = [\lambda(b)]^n \ . \tag{3.42}$$

Putting $\mu(u) = \ln \lambda(b)$ where $u = \ln b$, this criterion can be rewritten as

$$\mu(nu) = n\mu(u) \ . \tag{3.43}$$

Taking $u = 1$, we find that $\mu(n) = n\mu(1)$, i.e., that μ is a linear function of its argument. Hence,

$$\lambda(b) = b^y \tag{3.44}$$

where y is a scaling dimension that appears in the Widom scaling ansatz, and is fully specified by the eigenvalue λ_α and the scaling factor b. That is,

$$g(a_{\lambda_1}, a_{\lambda_2}, \ldots) = b^{-d} g(b^{y_1} a_{\lambda_1}, b^{y_2} \lambda_2 a_{\lambda_2}, \ldots) \tag{3.45}$$

The only difference between this and Widom's ansatz (2.33) is that the a_i are not necessarily the temperature or magnetic field, but combinations of these parameters. However, if one can be identified as a 'field-like' or a 'temperature-like' variable, we have a means to obtain thermodynamic critical exponents from the structure of the renormalisation group transformation.

3.7 A health warning about real-space renormalisation

We only have time to scratch the surface of the renormalisation group and, for pedagogical simplicity we have, along with everyone else, restricted ourselves to *real-space* renormalisation schemes. Although simple to visualise, these schemes are uncontrolled: there are many ways to cook up a transformation, and not all of them give the same results. It turns out that by transforming from real space into reciprocal (k) space, one can find a small parameter ϵ in which to perform a perturbative expansion. For Ising-like systems the ϵ in question is $4 - d$, i.e., the distance away from the upper critical dimension.

Exercises

3A: Renormalisation of the percolation transition

Site percolation is an example of a *geometric* phase transition. Initially, sites of a lattice are occupied randomly with probability p. If p is sufficiently large, it will be possible to get from one side of the lattice to the other using neighbouring occupied sites, while if p is too small, no connected path will exist. The first onset of a **spanning cluster** occurs at the percolation threshold p_c. Here we shall treat site percolation on a 2d triangular lattice via an approximate real-space renormalisation group technique.

(a) Coarse-grain the lattice by replacing each independent triangle of sites by a single site. This new site will assumed to be occupied if the majority of sites of the elementary triangle are occupied. Show that the renormalised occupation probability p' is given by

$$p' = p^3 + 3p^2(1-p) \ .$$

(b) Find the fixed points of this transformation and sketch the flow of the parameter p. Identify the values of p_c and the exponent ν defined by the divergence of a characteristic lengthscale $\xi \sim |p - p_c|^{-\nu}$. What does this lengthscale represent physically?

3B: Migdal-Kadanoff renormalisation of the 2d Ising model

A direct application of a Migdal-Kadanoff renormalisation scheme to the 2d Ising model leads to the proliferation of interactions discussed in the lecture. However, there is a trick that allows us to re-use the 1d scheme in a 2d model. We will consider a 2d square lattice where i and j label sites in the horizontal and vertical directions resepectively. The trick (which is an approximation) is to move bonds connected to sites with odd i and j so that all remaining bonds are 'doubled-up' (that is, have an associated interaction energy $2J$). The renormalised partition function Z' is achieved by summing over spins with (i, j) that are $(\text{odd}, \text{even})$ and $(\text{even}, \text{odd})$, so that all the remains is a sum over $(\text{even}, \text{even})$ spins.

(a) Show that the renormalised 2d partition function is the product of two renormalised 1d partition functions, and that the renormalised coupling constants transform as (3.20)–(3.22) but with K_1 replaced by $2K_1$ everywhere on the right-hand side. By considering the limits $K_1 \to 0$ and $K_1 \to \infty$, show that this apparently innocuous factor of 2 admits a non-trivial (repulsive) fixed point at $(K_1, K_2) = (K_c, 0)$ for some finite K_c.

(b) Obtain the eigenvectors and eigenvalues associated with the fixed point and sketch the renormalisation flow in the two-dimensional space of coupling constants. Show that the scaling dimensions associated with temperature and field respectively are $y_1 = 0.747$ and $y_2 = 0.879$ and, with the help of the results from Exercise 2A, state the values of the critical exponents $\alpha, \beta, \gamma, \delta$ and ν. (You may use the fact that $\tanh 4K_c \approx 0.8393$).

3C: Correlation functions in an extended mean-field theory

(a) Suppose that a correlation function $C(r)$ of a d-dimensional coordinate r satisfies the Green function equation

$$\left(\lambda^2 - \nabla^2\right) C(r) = A\delta(r)$$

where A is a constant (independent of r). Identify the characteristic lengthscale ξ of the correlation function and, using the form of the Laplacian appropriate for a system with d-dimensional spherical symmetry,

$$\nabla^2 = \frac{1}{r^{d-1}}\frac{\partial}{\partial r} r^{d-1}\frac{\partial}{\partial r} \ ,$$

where r is a scalar distance, show that the Green function equation can be rewritten as

$$\left(1 - \frac{1}{\rho^{d-1}}\frac{\partial}{\partial \rho}\rho^{d-1}\frac{\partial}{\partial \rho}\right) C(\rho) = a\delta(\rho)$$

where $\rho = r/\xi$ and the constant a are both dimensionless. Using the ansatz $C(\rho) = a\rho^{\gamma}K_{\mu}(\rho)$, show that for $\rho > 0$ and $d \geq 2$, the Green function must be a solution of Bessel's equation

$$z^2 K''_{\mu}(z) + zK'_{\mu}(z) - (z^2 + \mu^2)K_{\mu}$$

that vanishes at infinity, and identify the values of γ and μ.

The small- and large-argument approximations of the appropriate Bessel functions are

$$\begin{array}{lcll} K_{\mu}(z) & \sim & z^{-1/2}\mathrm{e}^{-z} & z \to \infty \\ K_{\mu}(z) & \sim & z^{-\mu} & z \to 0, \mu > 0 \\ K_0(z) & \sim & -\ln z & z \to 0 \end{array}$$

where, in all cases, we have dropped numerical prefactors. Use these to show that, in $d \geq 2$

$$C(r) \sim \xi^{2-d}\frac{\mathrm{e}^{-r/\xi}}{(r/\xi)^{\frac{d-1}{2}}}$$

when the correlation length ξ is finite (i.e., $T \neq T_c$). Show also that at the critical point, where $\xi \to \infty$, the ξ dependence vanishes leaving the pure power-law decay $C(r) \sim r^{-(d-2)}$.

(b) We will now derive the Green function equation for $C(r)$ within an extension to mean-field theory called **Ginzburg-Landau** theory so that the important parameter λ can be identified.

The starting point is a version of the Landau free energy defined for a spatially-varying order parameter field $\tilde{m}(x)$. Argue that to incorporate an energy cost associated with interfaces between regions of different order parameter, the leading term in a Landau-like expansion of the free energy is

$$\tilde{g}([\tilde{m}], T, h) = \int \mathrm{d}x \left[\frac{\tau}{2}\tilde{m}(x)^2 + \frac{b}{4}\tilde{m}(x)^4 - h(r)\tilde{m}(r) + \frac{\sigma}{2}\{\nabla\tilde{m}(x)\}^2\right] \ ,$$

where the notation $[\tilde{m}]$ is intended to indicate the dependence of $\tilde{g}$ on the *entire* function $\tilde{m}(x)$. [Hint: you may wish to look at the notes for Lecture 4]

For a given configuration of the order parameter field $\tilde{m}(x)$, the correlation function is related to the free energy functional by

$$C(x, x') = kT\frac{\delta\tilde{m}(x)}{\delta h(x')}$$

which involves the *functional derivative* defined as

$$\frac{\delta F[\tilde{m}(x)]}{\delta \tilde{m}(x')} = \lim_{\epsilon \to 0} \frac{F[\tilde{m}(x) + \epsilon \delta(x - x')] - F[\tilde{m}(x)]}{\epsilon}$$

and is a function of x'. Check the results

$$\begin{aligned} \frac{\delta}{\delta \tilde{m}(x')} \int \mathrm{d}x\, \tilde{m}(x)^k &= k \tilde{m}(x')^{k-1} \\ \frac{\delta}{\delta \tilde{m}(x')} \tilde{m}(x) &= \delta(x - x') \\ \frac{\delta}{\delta \tilde{m}(x')} \int \mathrm{d}x \frac{1}{2} [\nabla \tilde{m}(x)]^2 &= -\nabla^2 \tilde{m}(x') \ . \end{aligned}$$

You will need to integrate by parts and assume that surface terms vanish for the last part.

By minimising the Ginzburg-Landau free energy w.r.t. $\tilde{m}$ for some spatially-varying field $h(x)$ (i.e., taking the functional derivative and setting it to zero), obtain an equation of state for the spatially-varying order parameter field $\tilde{m}(x)$ in the form of a differential equation. Take the functional derivative of this equation of state w.r.t. h to obtain a Green function equation for the correlation function $C(r)$ where $r = |x - x'|$. Substitute in the equilibrium expressions for the order parameter in the absence of a field above and below T_c obtained from Landau theory to identify the parameter λ and therewith the exponent $\nu = \frac{1}{2}$ characterising the divergence of the correlation length ξ near criticality.

Lecture 4

Dynamic fluctuations at equilibrium

The second half is this course is devoted to more in-depth studies of dynamics. Starting from the principle of detailed balance derived in Lecture 1, we will derive an important stochastic equation of motion for a simple example system, the random **Brownian motion** of a single particle suspended in a fluid.

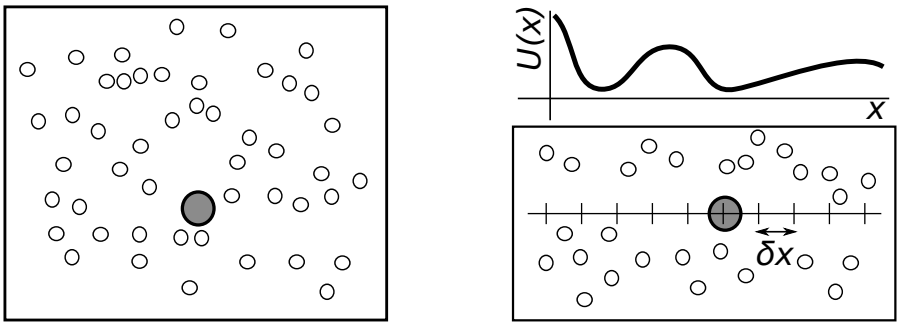
4.1 From detailed balance to a Fokker-Planck equation

We will go far by considering a very simple system—a single particle being thermally agitated by fluid particles in a heat bath—see Figure 4.1a. We will assume the degrees of freedom in the heat bath relax very quickly, so that the motion of the particle of interest is heavily retarded, or **overdamped**, by the bath. If this is the case, we expect to be able to model the motion of the particle as a Markov process (see Lecture 1).

The fundamental equation governing a Markov process is the master equation (see Lectures 1 and 2). For simplicity, let us initially assume that the particle is confined to sites of a one-dimensional lattice with small spacing δx as shown in Figure 4.1b. In the following, we will obtain a continuum description of the particle's motion by taking $\delta x \to 0$.

The master equation for this system reads

$$P(x,t+\delta t) = P(x-\delta x,t)P(x-\delta x\to x) + P(x+\delta x,t)P(x+\delta x\to x) + P(x,t)\left[1 - P(x\to x-\delta x) - P(x\to x+\delta x)\right] \quad (4.1)$$



(a) *A particle suspended in a fluid* **(b)** *Restriction to a 1d lattice*

Figure 4.1 *Simple system displaying stochastic dynamics: Brownian motion.*

where, to get the probability of being at x after one timestep, we sum over the probabilities of being on neighbouring sites at the previous timestep, and hopping onto x; and also the probability that we remain on site x. Rearranging

$$P(x,t+\delta t) - P(x,t) = P(x-\delta x,t)P(x-\delta x \to x) - P(x,t)P(x \to x-\delta x) + P(x+\delta x,t)P(x+\delta x \to x) - P(x,t)P(x \to x+\delta x)\ . \quad (4.2)$$

Let us assume the particle is sitting in a potential gradient that corresponds to a force pushing to the right. If we were to simulate the motion using the Metropolis algorithm (see Lecture 1), the transition probabilities would be

$$P(x \to x+\delta x) = 1 \qquad P(x \to x-\delta x) = \mathrm{e}^{-[U(x-\delta x)-U(x)]/kT} \quad (4.3)$$

$$P(x-\delta x \to x) = 1 \qquad P(x+\delta x \to x) = \mathrm{e}^{-[U(x)-U(x+\delta x)]/kT} \quad (4.4)$$

The idea now is to Taylor expand all the functions with arguments containing δx and δt:

$$P(x,t+\delta t) = P(x,t) + \dot{P}(x,t)\delta t + O([\delta t]^2) \quad (4.5)$$

$$P(x \pm \delta x,t) = P(x,t) + P'(x,t)\delta x + \frac{1}{2}P''(x,t)[\delta x]^2 + O([\delta x]^3) \quad (4.6)$$

$$P(x \to x-\delta x) = 1 + \frac{U'(x)}{kT}\delta x + \frac{1}{2}\left(\left[\frac{U'(x)}{kT}\right]^2 - \frac{U''(x)}{kT}\right)[\delta x]^2 + O([\delta x]^3) \quad (4.7)$$

$$P(x+\delta x \to x) = 1 + \frac{U'(x)}{kT}\delta x + \frac{1}{2}\left(\left[\frac{U'(x)}{kT}\right]^2 + \frac{U''(x)}{kT}\right)[\delta x]^2 + O([\delta x]^3)\ . \quad (4.8)$$

We see why we expanded to first order in time but second order in space when we substitute these expressions into the master equation—all the terms in δx cancel:

$$\dot{P}(x,t)\delta t + O([\delta t]^2) = \left[\frac{1}{kT}[U'(x)P(x,t)]' + \frac{1}{2}P''(x,t)\right][\delta x]^2 + O([\delta x]^3)\ . \quad (4.9)$$

We now want to take the continuous space and time limits, $\delta x \to 0$ and $\delta t \to 0$. If we take one limit before taking the other, all the terms simply vanish so this is clearly not the right thing to do. Instead, we take both δx and δt to zero simultaneously, holding the ratio $D = \frac{[\delta x]^2}{\delta t}$ fixed. Then:

$$\dot{P}(x,t) + O(\delta t) = D\left[\frac{1}{kT}[U'(x)P(x,t)]' + \frac{1}{2}P''(x,t)\right] + O([\delta t]^{1/2})\ . \quad (4.10)$$

Now, as $\delta t \to 0$, we obtain the differential equation that governs the evolution of the particle's probability distribution function:

$$\boxed{\frac{\partial}{\partial t}P(x,t) = -\frac{D}{kT}\frac{\partial}{\partial x}\left[-U'(x)P(x,t)\right] + D\frac{\partial^2}{\partial x^2}P(x,t)}\ . \quad (4.11)$$

This equation is called the **Fokker-Planck** equation and exactly describes the stochastic motion of an overdamped particle at equilibrium with a heat bath. Or at least it would, if we knew how to fix the kinetic parameter D.

4.2 Deterministic contribution to the dynamics

If the particle we're dealing with is large, we anticipate that the effect of the thermal fluctuations will be negligible. It will follow a path given by the equation of motion

$$m\ddot{x} + \gamma\dot{x} = -U'(x) \ . \tag{4.12}$$

To make contact with the Fokker-Planck equation, we first note that the overdamped limit corresponds to taking the size of the damping term relative to the inertial term to infinity: $\gamma/m \to \infty$. For a spherical particle of radius a in a fluid with viscosity η, the friction coefficient is given by Stokes' formula $\gamma = 6\pi a\eta$. The equation of motion for the particle is then

$$\dot{x} = -\frac{U'(x)}{6\pi a\eta} \ . \tag{4.13}$$

If thermal motion can be neglected, this coordinate $x(t)$ will correspond to the *mean* of the distribution $P(x,t)$. Multiplying both sides of (4.11) by x, integrating over all space, and taking $P(x,t) \to 0$ as $x \to \pm\infty$, we find that

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle x(t)\rangle = -D\frac{\langle U'(x[t])\rangle}{kT} \ . \tag{4.14}$$

Comparing (4.13) and (4.14) we conclude that the constant D is not free, but fixed by the deterministic behaviour of the particle:

$$D = \frac{kT}{\gamma} = \frac{kT}{6\pi a\eta} \ . \tag{4.15}$$

The first equality is sometimes called an **Einstein relation**, and the expression for D in terms of η a **Stokes-Einstein relation**. Both are examples of **fluctuation-dissipation relations**, of which we shall see more in due course.

4.3 Diffusive contribution to the dynamics: Brownian motion

To understand the physical meaning of the constant D, let us now switch off the external potential, so that the Fokker-Planck equation turns into the usual diffusion equation

$$\frac{\partial}{\partial t}P(x,t) = D\frac{\partial^2}{\partial x^2}P(x,t) \ . \tag{4.16}$$

The mean displacement of the particle is then zero, as can be seen from (4.13). The mean *square* displacement, meanwhile, grows without bound (at least, in an infinite system). This can be seen by multiplying the diffusion equation by x^2, integrating over all x by parts (twice). We find for a particle initially at the origin that

$$\langle x(t)^2\rangle = 2Dt \ , \tag{4.17}$$

an expression that defines the **diffusion constant** D. (Note that this is for just one particle coordinate; in a d-dimensional system, each of the d spatial coordinates diffuse independently, and one finds that $\langle x(t)\cdot x(t)\rangle = 2dDt$, for a particle that is initially at the origin).

The random motion of a particle in contact with a heat bath that ultimately gives rise to the linear growth of the mean square displacement had been observed by the botanist Robert Brown in 1827 when examining pollen grains under a microscope. Einstein was one of the first to recognise that the scale of the random fluctuations of the particle's position is necessarily related to the macroscopic frictional retardation of its velocity *by virtue of both originating from the same background fluid*. This stochastic motion also gives indirect evidence for the fluid to be composed of individual particles, rather than being some purely continuous medium.

4.4 Probability current and steady state

An appealing feature of the Fokker-Planck equation is that it can be couched in the form of a continuity equation

$$\frac{\partial}{\partial t}P(x,t) = -\frac{\partial}{\partial x}j(x,t) \tag{4.18}$$

because probability is conserved: the particle must be somewhere, so $\int \mathrm{d}x P(x,t) = 1$ at all times t. In this equation, $j(x,t)$ is the **probability current**

$$j(x,t) = -\frac{D}{kT}U'(x)P(x,t) - D\frac{\partial}{\partial x}P(x,t)\ . \tag{4.19}$$

Note that it depends on the probability distribution $P(x,t)$.

If left unperturbed, a finite system will eventually reach a **steady state**, i.e., one in which the probability distribution $P(x,t)$ reaches a time-independent limit $P^*(x)$. In this limit the current must approach a value that is constant over space. If the system is at equilibrium with the bath, this constant must equal zero, since there is no net flux of any quantity between the system and the bath. Putting $j(x,t) = 0$ in (4.19), and solving the resulting ordinary differential equation, we find, reassuringly, that

$$P^*(x) \propto \exp\left[-\frac{U(x)}{kT}\right]\ . \tag{4.20}$$

It is also possible to set up Fokker-Planck equations that describe nonequilibrium dynamics. In this case one usually finds a steady state for which $j(x,t) = j$, a nonzero constant, that must be determined from the dynamics. This is typically somewhat more difficult than finding the equilibrium current and corresponding steady-state distribution.

4.5 Alternative formulation of stochastic processes: Langevin equation

Langevin equations provide a way to write down stochastic dynamics that many physicists find attractive. The idea is to start off with a deterministic dynamics that you believe in, for example Eq. (4.12), and model the fluctuations in the particle velocity due to collisions with particles in the heat bath by means of a **rapidly fluctuating force** $\xi(t)$, also called **noise**. So, for over-damped motion of a single particle in a fluid, one would write

$$\gamma\dot{x}(t) = f(x) + \xi(t)\ . \tag{4.21}$$

This 'just add noise' prescription is very widespread, and can be extended to many-particle dynamics as we will see below.

It is important to realise that in the Langevin equation (4.21) a specific *realisation* of the noise $\xi(t)$ is implied: that is, like any other function it has a well-defined value at any given time. The point is that this value is different each time you rerun the process past the same time instant. We are usually interested in averages of the variable $x(t)$, which is a random function of time because $\xi(t)$ is a random function of time.

The mean value of the noise $\xi(t)$ is easy to work out: it must be zero so that when we average (4.21) we obtain the desired deterministic equation (4.14). We also assume that the noise is so rapidly fluctuating that correlations between its values at two different times decay very quickly. This is justified for a Brownian particle, for example, because whilst its velocity is correlated on a timescale of order

of milliseconds, collisions with fluid particles occur roughly every 10^{-14} seconds! We thus idealise the noise by saying that its values at any two (different) times are totally uncorrelated. That is, we write

$$\langle \xi(t) \rangle = 0 \quad \text{and} \quad \langle \xi(t)\xi(t') \rangle = \Gamma \delta(t - t') \ . \tag{4.22}$$

Why is $\langle \xi(t)\xi(t') \rangle$ not uniformly zero? The answer is that any random variable is always correlated with itself: by putting in the delta function we can also state the variance of the noise distribution as Γ. These two constraints are not sufficient to specify the full noise distribution: it is usually further stated that the noise has a Gaussian distribution, which is completely specified by its mean and variance. Again, by considering the large number of fluid particles that Brownian particle may come into contact with, we can justify a Gaussian distribution of the forces exerted on the Brownian particle by appealing to the central limit theorem.

Consider once again the case of an unforced particle, $f(x) = 0$. We can integrate (4.21) for a given realisation $\xi(t)$ to find

$$x(t) = x(0) + \frac{1}{\gamma} \int_0^t \mathrm{d}t' \xi(t') \ . \tag{4.23}$$

The mean displacement of the particle is

$$\langle [x(t) - x(0)] \rangle = \frac{1}{\gamma} \int_0^t \mathrm{d}t' \langle \xi(t') \rangle = 0 \ , \tag{4.24}$$

as we previously found. Its mean square displacement is

$$\langle [x(t) - x(0)]^2 \rangle = \frac{1}{\gamma^2} \int_0^t \mathrm{d}t_1 \int_0^t \mathrm{d}t_2 \langle \xi(t_1)\xi(t_2) \rangle = \frac{\Gamma t}{\gamma^2} \ . \tag{4.25}$$

Comparing with (4.17) we can state that the noise strength Γ can be fixed in terms of other quantities

$$\Gamma = 2\gamma^2 D = 2\gamma kT \ . \tag{4.26}$$

This is another fluctuation-dissipation relation, since it relates Γ, a measure of the size of the fluctuations, to γ, a measure of the resistance felt by the Brownian particle. In the next lecture, we will show that the Fokker-Planck and Langevin equations are equivalent.

4.6* Coarse-grained many-body dynamics

Langevin equations are used extensively in statistical mechanics to model temporal fluctuations. In particular, they can be applied to spontaneously ordering magnetic or fluid systems to establish properties of the dynamics at and near the critical point. These dynamics are often couched in terms of some coarse-grained order-parameter field, for example, the magnetisation averaged over microscopically large, but macroscopically small, regions of space. Let $\tilde{m}_x$ be the order-parameter density around some point x. The dynamics of this order parameter field should also be governed by the detailed balance principle. Hence, we should ultimately obtain a Langevin equation

$$\frac{\mathrm{d}}{\mathrm{d}t} \tilde{m}_x = -\frac{\partial}{\partial \tilde{m}_x} g(\tilde{\mathbf{m}}_x, T, h) + \xi_x(t) \tag{4.27}$$

where the strength of the noise is fixed by the fluctuation-dissipation relation.

In principle, the Gibbs free energy density depends on *all* the coarse-grained local magnetisation variables $\tilde{m}_x$. In the continuum limit, this many-variable function turns into a functional

$$g([\tilde{m}], T, h) = \int \mathrm{d}x g_\ell(\tilde{m}(x), T, h) \tag{4.28}$$

where $g_\ell(m(x), T, h)$ is a local free energy density. The corresponding Langevin equation then involves a **functional derivative**

$$\frac{\partial}{\partial t}\tilde{m}(x) = -\frac{\delta}{\delta\tilde{m}(x)}g([\tilde{m}], T, h) + \xi(x,t) \tag{4.29}$$

and the noise is a function both of space and time. The functional derivative is formally defined through

$$F[y(x) + \epsilon\delta(x - x_0)] = F[y(x)] + \epsilon\frac{\delta F[y(x)]}{\delta y(x_0)} + O(\epsilon^2) \ . \tag{4.30}$$

For simplicity, the noise is assumed to exhibit neither spatial nor temporal correlations:

$$\langle\xi(x,t)\xi(x',t')\rangle = \Gamma\delta(x - x')\delta(t - t') \ . \tag{4.31}$$

This coarse-grained formulation of the stochastic dynamics admits the total magnetisation in the system to fluctuate: it was derived from the Gibbs free energy. It is known as **Model A** in a classification scheme ascribed to Hohenberg and Halperin (1977).

Various other models are available—the scheme goes up to at least Model H. The only other one we will consider is **Model B**, which concerns the dynamics when the order parameter is conserved. In principle one would obtain Model B by projecting the Model A dynamics onto the subspace in which the order parameter is conserved. In practice, we observe that conservation is guaranteed if the dynamics can be expressed in terms of a current $j(x,t)$ via

$$\frac{\partial}{\partial t}\tilde{m}(x) = -\nabla\cdot j(x,t) \ . \tag{4.32}$$

To find this current we note that the derivative of the free energy w.r.t. the order parameter is a field, and that the (negative) gradient of this field defines the direction of the systematic part of the current:

$$j(x,t) = -\nabla\frac{\delta}{\delta\tilde{m}(x)}g([\tilde{m}], T, h) \ . \tag{4.33}$$

To this current we add noise (to each component of j independently), and using (4.32) obtain the Model B dynamics

$$\frac{\partial}{\partial t}\tilde{m}(x) = \nabla^2\frac{\delta}{\delta\tilde{m}(x)}g([\tilde{m}], T, h) + \zeta(x,t) \tag{4.34}$$

where the 'conserved' noise has zero mean and a correlator

$$\langle\zeta(x,t)\zeta(x',t')\rangle = -\Gamma\nabla^2\delta(x - x')\delta(t - t') \ . \tag{4.35}$$

4.7* Critical slowing down

We can use Models A and B to examine the relaxation of fluctuations to equilibrium near a critical point. To do this we need to write down a version of the Landau free energy density that admits spatial variation in the order parameter. In practice, this means including gradient terms in the local free energy density and, in the spirit of Landau theory, discarding terms incompatible with the symmetries of the system and keeping only the leading terms. The resulting prescription is

$$g_\ell(\tilde{m}(x), T, h) = \frac{a}{2}\tilde{m}^2 + \frac{b}{4}\tilde{m}^4 - h\tilde{m} + \frac{\sigma}{2}(\nabla\tilde{m})^2 \ . \tag{4.36}$$

The first three terms we recognise from the Landau theory of Lecture 2, (2.29). Recall that the parameter $a \sim T - T_c$ so that the ordered phase is stable below T_c. For simplicity, we set $a = \tau = \frac{T-T_c}{T_c}$ in the following. The final term imposes a free energy penalty for gradients of the order parameter: these

are associated with interfaces between domains of different magnetisation. Since the orientation of the interface does not affect the free energy, the lowest power of the gradient that may contribute is 2.

We can use this free energy density to gain an insight into the dynamics near the critical point. For our purposes, we can look at the approach from the high-temperature phase where the minimum of the free energy is around $\tilde{m} = 0$. This allows us to dispense with the quartic term (which serves only to stabilise the system for large $\tilde{m}$). The resulting dynamics, can be written in a unified form for both Models A and B as

$$\frac{\partial \tilde{m}}{\partial t} = -(i\nabla)^s \left[(\tau - \nabla^2)\tilde{m}(x,t) - h(x,t) \right] + \zeta_s(x,t) \tag{4.37}$$

where the noise has zero mean and correlator

$$\langle \zeta_s(x,t)\zeta_s(x',t') \rangle = \Gamma(i\nabla)^s \delta(x-x')\delta(t-t') \ . \tag{4.38}$$

With the parameter set to $s = 0$, we recover Model A; and with $s = 2$ we get Model B.

Suppose that, during relaxation from a perturbation around $\tilde{m} = 0$ in zero field, the order parameter field has the scaling form $\tilde{m}(x,t) = \tilde{m}(x/\xi, t/\theta)$. Substituting into (4.37), and ignoring the noise, we find

$$\frac{1}{\theta}\frac{\partial \tilde{m}}{\partial t} = -i^s \frac{1}{\xi^s}\nabla^s \left(\tau - \frac{1}{\xi^2}\nabla^2 \right) \tilde{m} \ . \tag{4.39}$$

Since $\xi \sim |\tau|^{-\nu}$ with $\nu = \frac{1}{2}$ for this model (see Lecture 3), we have that the right-hand side scales as $\xi^{-(2+s)}$. Meanwhile the left-hand side scales as θ^{-1}. For the scaling ansatz to be self-consistent we require that the characteristic time and length scales are related by

$$\boxed{\theta \sim \xi^z} \tag{4.40}$$

where z defines a new critical exponent, the **dynamical exponent**, which, in mean field theory, takes the values

$$z = \begin{cases} 2 & \text{Model A} \\ 4 & \text{Model B} \end{cases} \tag{4.41}$$

The point of doing this analysis is to demonstrate: (i) symmetries and conservation laws define *dynamic* universality classes in addition to static ones; (ii) one static universality class may be associated with multiple dynamics universality classes; (iii) near a critical point, one has a diverging correlation time in addition to a diverging correlation length. This last fact is of particular relevance in computer simulation: in order to get good statistics, one needs system sizes to be at least a few correlation lengths and to run for at least a few correlation times. Since both diverge near criticality, simulations of critical phenomena can be very difficult to perform. It is, however, sometimes possible to design the dynamics in such a way to reduce the dynamic exponent z close to zero in order to alleviate this problem; the Wolff cluster algorithm discussed in Exercise 1C provides one example.

Exercises

4A: Continuum limit of a random walk

An unbiased random walk is a process in which a step of size δx is taken in each timestep δt to the left and to the right with probability $\frac{1}{2}$.

(a) Assuming the walker starts at $x = 0$ at time $t = 0$, write down the probability distribution $P(x,t)$. [Hint: this is essentially the same question as 1A].

(b) For $x/\delta x$ and $t/\delta t$ both large, show that the peak of the distribution can be approximated as a Gaussian

$$P(x,t) \sim \frac{1}{Z(t)} \mathrm{e}^{-x^2/4Dt}$$

for a suitable choice of the diffusion constant D. Calculate the normalisation $Z(t)$.

(c) Obtain the jump moments for the process and hence show that the continuum limit of the random walk is pure diffusion. Verify that the Gaussian distribution obtained above is the solution of the resulting diffusion equation subject to the initial condition $P(x,0) = \delta(x)$.

4B: Correlation functions and Langevin equations

(a) Give a physical interpretation of the Langevin equation

$$m\dot{v} = -\gamma v + \xi \quad \text{where} \quad \langle \xi(t)\xi(t') \rangle = \Gamma \delta(t-t') \ .$$

Do we expect the variance of the noise Γ for this process to be the same as that described by (4.21)?

(b) Calculate the velocity autocorrelation function

$$\langle v(t_1)v(t_2) \rangle = \left(v(0)^2 - \frac{\Gamma}{2\gamma m} \right) \mathrm{e}^{-\gamma(t_1+t_2)/m} + \frac{\Gamma}{2\gamma m} \mathrm{e}^{-\gamma|t_2-t_1|/m}$$

and use this to establish the noise strength Γ. Compare the result with that obtained for over-damped motion (4.26).

4C: Model A dynamics in the 1d Ising model at $T = 0$

In one spatial dimension and the limit of zero temperature, a dynamics for the Ising model in which single spins are flipped with a probability given by the heat-bath algorithm defined in Exercise 1B can be implemented as follows:

1. Pick a pair of neighbouring sites (i,j) at random, where $j = i \pm 1$.
2. If spins s_i and s_j are not aligned, flip spin j.
3. Go back to step 1, stopping once all spins are aligned.

Explain why this is the case, and why this is a realisation of Model A dynamics.

Show that, as a consequence of the master equation (2.4), the change in the average of an observable $A(\mathbf{s})$ per time step δt is

$$\frac{\delta \langle A(\mathbf{s}) \rangle}{\delta t} = \sum_{i=1}^{N} \langle P_{\delta t}(s_i \to -s_i)[A(s_1, \cdots, -s_i, \cdots, s_N) - A(\mathbf{s})] \rangle$$

where, for zero-temperature heat-bath dynamics,

$$P(s_i \to -s_i) = \frac{1}{4N}(2 - s_i s_{i-1} - s_i s_{i+1})$$

and N is the number of spins of a 1d periodic system ($s_i = s_{i+N}$).

Defining $g_k(t) = \frac{1}{N} \sum_i \langle s_i s_{i+k} \rangle$, show that the mean length of a domain of spins $\ell = 2/(1-g_1)$, and that for $k \geq 1$,

$$\frac{\mathrm{d}g_k}{\mathrm{d}t} = g_{k-1}(t) - 2g_k(t) + g_{k+1}(t)$$

in a suitable choice of time units, subject to the boundary condition $g_0(t) = 1$.

Construct the generating function $F(z,t) = \sum_{k=-\infty}^{\infty} g_k(t) z^k$ and show that it must satisfy

$$\frac{\partial}{\partial t} F(z,t) = \left(z + \frac{1}{z} - 2 \right) F(z,t) \ .$$

Solve this differential equation, and show that the solution with initial condition $g_k(0) = \delta_{k,k_0}$ is $g_k(t) = \mathrm{e}^{-2t} I_{k-k_0}(2t)$ where $I_n(u)$ is a modified Bessel function of the first kind defined by its generating function

$$\sum_{n=-\infty}^{\infty} I_n(u) z^n = \exp\left[\frac{(z + z^{-1})u}{2} \right]$$

and behaves asymptotically for large argument as

$$I_n(u) = \frac{\mathrm{e}^u}{\sqrt{2\pi u}} + O(u^{-3/2}) \ .$$

Now show that the particular linear combination of Bessel functions

$$g_k(t) = 1 - \mathrm{e}^{-2t} \sum_{n=1}^{\infty} \left[I_{k-n}(2t) - I_{k+n}(2t) \right]$$

is the one that satisfies the boundary condition $g_0(t) = 1$ for all t and an infinite-temperature initial condition, i.e., one in which all spins point up or down with equal probability. [Hint: $I_{-n}(u) = I_n(u)$ and $I_n(0) = \delta_{n,0}$.] Hence, show that

$$\ell = \frac{2\mathrm{e}^{2t}}{I_0(2t) + I_1(2t)} \sim 2\sqrt{\pi t}$$

This provides an example of a **coarsening process** characterised by the coarsening exponent $\frac{1}{2}$. A more general analysis of these processes shows that this power-law growth is a generic property of Model A dynamics following a quench to any temperature below T_c, not just to zero temperature. A good review of coarsening processes is by Bray (1994).

Lecture 5

Near-equilibrium fluctuations

In these last two lectures we consider progressively further-from-equilibrium phenomena: that is, what happens if we perturb the dynamics slightly (or even a lot—see next lecture). We begin by using the detailed balance relation to understand how an observable quantity responds to the application of a small amount of work on the system. Then, we will systemise the derivation of the Fokker-Planck equation that was given in the last lecture, and use this to understand what happens when we modify the dynamics of a system at its boundaries (e.g., by allowing particles to stick to the walls of a container).

5.1 Fluctuation-dissipation relation

In the last lecture we saw two examples of a **fluctuation-dissipation relation**. The first gave an expression for the diffusion constant of a particle, D, in terms of friction coefficient of the surrounding medium, γ, via $D = \frac{kT}{\gamma}$. The second related the amplitude, Γ, of the stochastic Langevin force to the friction coefficient, γ, via $\Gamma = 2\gamma kT$. These arose in the special case of a particle diffusing in a one-dimensional potential. Both are in fact a special case of a more general fluctuation dissipation relation which is stated as follows.

A system is prepared in its equilibrium state, i.e., with its microstate X drawn from the Boltzmann distribution $\mathrm{e}^{-U(X)/kT}/Z$, and then evolves under usual equilibrium dynamics up to time $t = 0$. At which point a small amount of work ΔW is done on the system. Thereafter, the system continues to evolve under its usual equilibrium dynamics up to some later time t, at which point the system arrives in the microstate state X_t, and the quantity $B(t) \equiv B(X_t)$ is observed.

To specify the fluctuation-dissipation relation, we need to be more precise about *how* work is done. For work that is done over a time interval $[0, \delta t]$, we have

$$\Delta W = f\Delta A(X) = f[A(X_{\delta t}) - A(X_0)] \tag{5.1}$$

where f is a force and A some quantity conjugate to that force. (For example, if f is a force acting on a particle, $A(X)$ would be the position of that particle). Now, taking $\delta t \to 0$, we have

$$\Delta W = \lim_{\delta t \to 0} f \frac{\mathrm{d}A(X_t)}{\mathrm{d}t} \delta t = h\dot{A}(t) \tag{5.2}$$

where h is the impulse

$$h = \lim_{\delta t \to 0} f \delta t \tag{5.3}$$

which is assumed to be finite, and $A(t) \equiv A(X_t)$.

The fluctuation-dissipation relation then tells us how the average $\langle B(t)\rangle$ changes under the application of the work ΔW for small impulses h. Specifically, it says that

$$\frac{\mathrm{d}}{\mathrm{d}h}\langle B(t)\rangle_h\bigg|_{h=0} = \frac{1}{kT}\langle \dot{A}(0)B(t)\rangle_0 \ . \tag{5.4}$$

The significance of this relation is that the dynamical ensembles over which averages are taken are different on the two sides. On the left-hand side, the notation $\langle\cdot\rangle_h$ denotes an average over a set of trajectories where an external perturbation with impulse h is applied at time $t = 0$. That is, there is some (albeit small) interaction between the observer and the system of interest in this ensemble. On the right-hand size, the notation $\langle\cdot\rangle_0$ denotes an average over a set of trajectories where no such perturbation is applied: i.e., the system of interest is just observed in its equilibrium state. This means, that *it is sufficient to observe the equilibrium fluctuations of a system to learn about its response to an applied force* (at least if that force is small).

We will see the utility of this shortly, but first let us show that the relation $D = \frac{kT}{\gamma}$ is a special case, as previously claimed. To do this, we note that the expression (4.17) for the diffusion constant D can be written as

$$2D = \lim_{\delta t\to 0}\frac{\langle[x(\delta t) - x(0)]^2\rangle_0}{\delta t} \tag{5.5}$$

where we have taken the time interval to be δt and the initial particle position to be $x(0)$. Now, if the particle is at equilibrium, then $\langle x(\delta t)^2\rangle = \langle x(0)^2\rangle$ and

$$2D = \lim_{\delta t\to 0}\frac{\langle[x(\delta t)^2 - 2x(\delta t)x(0) + x(0)^2]\rangle_0}{\delta t} = 2\lim_{\delta t\to 0}\frac{\langle x(\delta t)[x(\delta t) - x(0)]\rangle_0}{\delta t} \tag{5.6}$$

$$\implies D = \langle \dot{x}(0)x(0^+)\rangle_0 \tag{5.7}$$

where $x(0^+)$ is the position of the particle just after the force is applied. We see that this has the form of the right-hand side of (5.4) if we identify both A and B with the particle position x. The left-hand side meanwhile can be written as

$$\frac{\mathrm{d}}{\mathrm{d}h}\langle x(0^+)\rangle_h\bigg|_{h=0} = \lim_{h\to 0}\frac{\langle x(0^+)\rangle_h - \langle x(0^-)\rangle_0}{h} = \lim_{h\to 0}\frac{\langle[x(0^+) - x(0^-)]\rangle_h}{h} \tag{5.8}$$

where we have used the fact that $\langle x(0^+)\rangle_0 = \langle x(0^-)\rangle_0$, i.e., that the mean position of the particle is constant at equilibrium. We have also used the fact that before the perturbation is applied (i.e., at time 0^-), the perturbed and unperturbed ensembles are the same: $\langle x(0^-)\rangle_h = \langle x(0^-)\rangle_0$. Now, the Langevin equation for the overdamped particle with an impulse of size h applied at time $t = 0$ is

$$\gamma\dot{x}(t) = -\frac{\mathrm{d}V(x)}{\mathrm{d}x} + h\delta(t) + \eta(t) \ . \tag{5.9}$$

Integrating this from $t = 0^-$ to $t = 0^+$, and taking the average, we find that

$$\lim_{h\to 0}\frac{\langle[x(0^+) - x(0^-)]\rangle_h}{h} = \frac{1}{\gamma} \ . \tag{5.10}$$

Now we have both sides of (5.4), we finally arrive at $D = \frac{kT}{\gamma}$, as required.

Note that we never said that starting from (5.4) we would arrive more quickly at earlier results than by other paths; just that (5.4) is more general. We will see below that the fluctuation-dissipation relation (5.4) is a consequence of the detailed balance relation, which applies to equilibrium systems of arbitrary

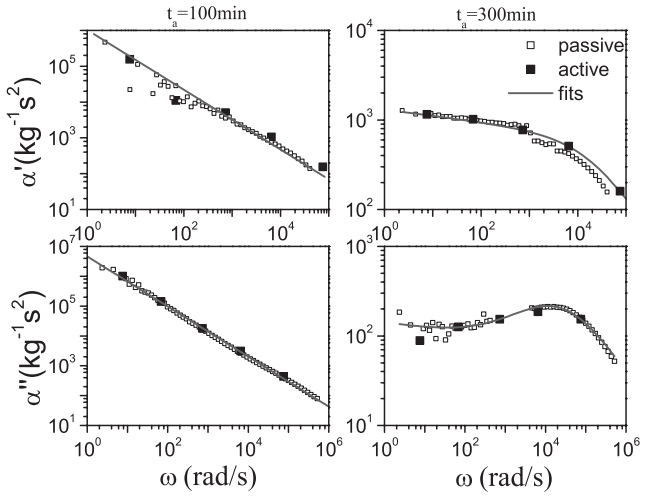


Figure 5.1 *Comparison of direct ('active') and indirect ('passive'), via the fluctuation-dissipation relation) measurements of response functions α' and α'' in a dense colloidal system. Figure from Jabbari-Farouji et al. (2007).*

complexity. Therefore we are not limited to single-particle systems in applications, and it has indeed been used, for example, to determine response functions of a dense colloidal system without applying any external forces. A comparison of these *passive* measurements (obtained by observing fluctuations at equilibrium) with their *active* counterparts (obtained by forcing the system) is show in Figure 5.1. Study the figure closely and eventually you will see that there are two data sets present, one on top of the other. This provides direct confirmation of the validity of the fluctuation-dissipation relation. The reason these data appeared in Physical Review Letters is because it had previously been thought that a dense colloidal 'glass' should be representative of systems that are not at equilibrium and display **ageing** behaviour. Therefore one does not expect the fluctuation-dissipation relation to hold in such systems. Note that it turns out to be more straightforward experimentally to work in the frequency rather than the time domain, that is, with the Fourier transformed version of (5.4).

5.2 Proof of the fluctuation-dissipation relation

Right at the start of the course, we showed that the probability of a given trajectory, X, of a system of interest satisfies the detailed balance principle

$$P[X|X_0] = P[\hat{X}|\hat{X}_0]\mathrm{e}^{\Delta Q[X]/kT} , \tag{5.11}$$

where $\hat{X}$ is the time-reversed trajectory, and $\Delta Q[X]$ is the heat exported to the surrounding reservoir along the trajectory X. This principle applies when the system of interest is in equilibrium with the reservoir. The basic physical assumption that underpins the fluctuation-dissipation relation is **Onsager's regression hypothesis** (Onsager, 1931a,b). This states that any change to the system of interest arising from a small external perturbation could in principle also arise spontaneously through an equilibrium fluctuation. The system of interest does not know how such a change was generated, and therefore the probability of following the trajectory $P[X|X_0]$ should be the same for both types of initial condition, and in particular satisfy the detailed balance principle.

To make use of it, we note that the first law implies that

$$\Delta Q[X] = U(X_{0^-}) - U(X_t) + \Delta W \tag{5.12}$$

where $U(X_{0-})$ is the internal energy of the system of interest before the perturbation is applied. Thus, we can rewrite (5.11) as

$$\frac{\mathrm{e}^{-U(X_{0-})/kT}}{Z}P[X|X_0]\mathrm{e}^{-\Delta W/kT} = \frac{\mathrm{e}^{-U(X_t)/kT}}{Z}P[\hat{X}|\hat{X}_0] \ . \tag{5.13}$$

Now we multiply both sides by some function $B(X_t)$ of the end-point of the forward trajectory (which is the start point of the reverse trajectory), integrate over all trajectories,

$$\int \mathrm{d}X \frac{\mathrm{e}^{-U(X_{0-})/kT}}{Z}P[X|X_0]\mathrm{e}^{-\Delta W/kT}B(X_t) = \int \mathrm{d}X \frac{\mathrm{e}^{-U(X_t)/kT}}{Z}P[\hat{X}|\hat{X}_0]B(X_t) \ , \tag{5.14}$$

and think hard. The right-hand side involves an integral of $B(X_t)$ over the equilibrium distribution of the final point of the trajectory X_t, weighted then by an integral over all paths starting from that final point (going backwards in time). This latter integral is just 1: the total probability of all paths must be properly normalised. So the right-hand side is just the equilibrium average $\langle B(X_t)\rangle_0 \equiv \langle B(t)\rangle_0$. The left-hand side, however, is an integral over the *perturbed* ensemble. We can therefore write the previous equation more concisely as

$$\left\langle \mathrm{e}^{-\Delta W/kT}B(t)\right\rangle_h = \langle B(t)\rangle_0 \ . \tag{5.15}$$

Now we recall that $\Delta W = h\dot{A}(0)$, and expand to first in powers of h. We then find that

$$\langle B(t)\rangle_h - \frac{h}{kT}\langle \dot{A}(0)B(t)\rangle_h + O(h^2) = \langle B(t)\rangle_0 \ . \tag{5.16}$$

Now, we can also expand the second average in powers of h as

$$\langle \dot{A}(0)B(t)\rangle_h = \langle \dot{A}(0)B(t)\rangle_0 + O(h) \tag{5.17}$$

and thus we find that

$$\frac{\langle B(t)\rangle_h - \langle B(t)\rangle_0}{h} = \frac{1}{kT}\langle \dot{A}(0)B(t)\rangle_0 + O(h) \ , \tag{5.18}$$

which is exactly equivalent to (5.4).

The eagle-eyed will have noticed that we obtained (5.15) without making the assumption that h was sufficiently small that nonlinear terms could be neglected. This observation opens the door to understanding far from equilibrium processes, as we will see in the next lecture.

5.3 Kramers-Moyal expansion of the Fokker-Planck equation

We now shift focus to a different class of perturbation of the equilibrium dynamics that comprises those that are induced by changing the behaviour of the system at its boundaries. To handle such perturbations, it is helpful to establish the general formulation of a stochastic process as a Fokker-Planck equation. As we will see, this brings us some further benefits such as proving the equivalence of the Fokker-Planck and Langevin equations and the ability to construct models of far-from-equilibrium dynamics.

We begin with the joint probability distribution $P(x,t|x_0,t_0)$ that a particle found at x_0 at some time t_0 is at the point x at a later time t. (Strictly, $P(x,t|x_0,t_0)\mathrm{d}x$ gives the probability of finding the particle in the infinitesimal interval $[x, x+\mathrm{d}x]$, which is what is implied by the word 'at'). If a process is Markov, this distribution does not depend on the trajectory that particle has followed at times earlier than t_0. Put another way, the transitions $(x_0,t_0) \to (x_1,t_1)$ and $(x_1,t_1) \to (x_2,t_2)$ are independent. Thus we can write the decomposition

$$P(x_2,t_2|x_0,t_0) = \int \mathrm{d}x_1 P(x_2,t_2|x_1,t_1)P(x_1,t_1|x_0,t_0) \ . \tag{5.19}$$

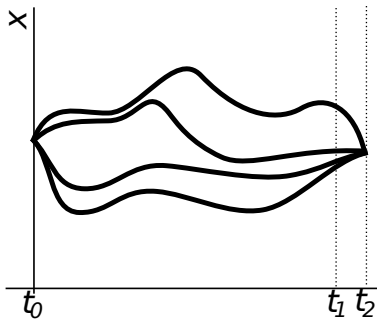


Figure 5.2 *Realisations of a stochastic process passing through (x_0, t_0) and (x_2, t_2). Taking a small interval at the end of the process and Taylor expanding leads ultimately to the Fokker-Planck equation.*

(In reality, this is just another version of the master equation, Eq. (2.4)). The idea now is to convert this integral equation into a differential equation by making one of the time intervals $t_1 - t_0$ or $t_2 - t_1$ infinitesimal, i.e., equal to $\mathrm{d}t$. We'll take this to be the second interval, as shown in Figure 5.2.

To keep track of spatial derivatives, we multiply both sides by an arbitrary test function $f(x_2)$ and integrate over all x_2:

$$\int \mathrm{d}x_2 f(x_2) P(x_2, t_1 + \mathrm{d}t | x_0, t_0) = \int \mathrm{d}x_2 f(x_2) \int \mathrm{d}x_1 P(x_2, t_1 + \mathrm{d}t | x_1, t_1) P(x_1, t_1 | x_0, t_0) \ . \tag{5.20}$$

Expanding the left-hand side in powers of $\mathrm{d}t$ we get

$$\text{l.h.s.} = \int \mathrm{d}x_1 f(x_1) \left[P(x_1, t_1 | x_0, t_0) + \frac{\partial}{\partial t_1} P(x_1, t_1 | x_0, t_0) \mathrm{d}t \right] + O(\mathrm{d}t^2) \tag{5.21}$$

after relabelling x_2 as x_1 for almost immediate convenience. On the right-hand side the transition $(x_1, t_1) \to (x_2, t_1 + \mathrm{d}t)$ takes only a short time $\mathrm{d}t$, and so we anticipate that the displacement $x_2 - x_1$ will also be small. Therefore, we perform the **Kramers-Moyal** expansion of the test function about x_1 in powers of this displacement:

$$\text{r.h.s.} = \int \mathrm{d}x_1 \int \mathrm{d}x_2 \left[\sum_{n=0}^{\infty} \frac{1}{n!} \frac{\mathrm{d}^n f(x_1)}{\mathrm{d}x_1^n} (x_2 - x_1)^n \right] P(x_2, t_1 + \mathrm{d}t | x_1, t_1) P(x_1, t_1 | x_0, t_0) \ . \tag{5.22}$$

To simplify this expression we introduce the **jump moments** $a_n(x, t)$ through

$$\langle [x(t + \mathrm{d}t) - x(t)]^n \rangle \equiv \int \mathrm{d}x' (x' - x)^n P(x', t + \mathrm{d}t | x, t) = a_n(x, t) \mathrm{d}t + O(\mathrm{d}t^2) \quad n \geq 1 \ . \tag{5.23}$$

These quantities give some measure of how far the particle moves in an infinitesimal time-step starting at (x, t). We now find that

$$\text{r.h.s.} = \int \mathrm{d}x_1 \left[f(x_1) P(x_1, t_1 | x_0, t_0) + \sum_{n=1}^{\infty} \frac{a_n(x_1, t_1)}{n!} \frac{\mathrm{d}^n f(x_1)}{\mathrm{d}x_1^n} P(x_1, t_1 | x_0, t_0) \mathrm{d}t \right] + O(\mathrm{d}t^2) \ . \tag{5.24}$$

The term of order 1 (in powers of $\mathrm{d}t$) on the l.h.s. and r.h.s. cancel; to handle the terms of order $\mathrm{d}t$, we integrate the r.h.s. by parts and, assuming f, P and all their derivatives vanish at infinity, we find

$$0 = \int \mathrm{d}x_1 f(x_1) \left[\frac{\partial}{\partial t_1} - \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x_1^n} a_n(x_1, t_1) \right] P(x_1, t_1 | x_0, t_0) + O(\mathrm{d}t) \tag{5.25}$$

Taking the limit $\mathrm{d}t \to 0$ and insisting that this equation hold for any test function $f(x)$, we obtain the equation

$$\boxed{\begin{aligned}\frac{\partial}{\partial t_1}P(x_1,t_1|x_0,t_0) &= \sum_{n=1}^{\infty}\frac{(-1)^n}{n!}\frac{\partial^n}{\partial x_1^n}a_n(x_1,t_1)P(x_1,t_1|x_0,t_0)\\ &\text{where}\\ a_n(x,t) &= \lim_{\delta t\to 0}\frac{\langle[x(t+\mathrm{d}t)-x(t)]^n\rangle}{\mathrm{d}t}\ .\end{aligned}} \tag{5.26}$$

This equation is called the **forward Fokker-Planck** (or sometimes **Kolmogorov**) equation.

5.4 At what order do we stop in the expansion?

Having gone to the effort of deriving this equation, what do we now do with it? The utility of the Kramers-Moyal expansion is that one can formulate a Fokker-Planck equation for any stochastic process governing a continuous variable. All one needs to do is calculate the jump moments. One may ask whether we need to calculate the entire (infinite) set of moments. In principle, yes. However, a theorem (due to Pawula) tells us that the series terminates either after the first time ($n = 1$), the second ($n = 2$), or does not terminate. Furthermore, if *any* of the even jump moments vanish, all a_n with $n > 2$ vanish. Therefore, it is most common to see only the first two terms written down; and indeed, even when the expansion does not terminate, one often gets a reasonable approximation to the dynamics by just keeping the first two terms.

5.5 Equivalence of the Fokker-Planck and Langevin equations

One thing we can do with the general form of the Fokker-Planck equation is to show that the Langevin equation

$$\gamma\dot{x}(t) = -\frac{\mathrm{d}}{\mathrm{d}x}U(x) + \xi(t) \tag{5.27}$$

where

$$\langle\xi(t)\rangle = 0 \quad \text{and} \quad \langle\xi(t)\xi(t')\rangle = \Gamma\delta(t-t') \tag{5.28}$$

is equivalent to the Fokker-Planck equation

$$\frac{\partial}{\partial t}P(x,t) = -\frac{D}{kT}\frac{\partial}{\partial x}\left[-U'(x)P(x,t)\right] + D\frac{\partial^2}{\partial x^2}P(x,t)\ , \tag{5.29}$$

with the noise strength Γ related to the diffusion constant and friction coefficient via $\Gamma = 2\gamma^2 D$, as was claimed in the last lecture. To achieve this we must calculate the jump moments $a_n(x,t)$ from the Langevin equation (5.27). The key relation is

$$x(t+\mathrm{d}t) - x(t) = \frac{1}{\gamma}\int_t^{t+\mathrm{d}t}\mathrm{d}t'\left\{f(x[t']) + \xi(t')\right\}\ . \tag{5.30}$$

Taking the n^{th} power, and averaging over the Gaussian distribution of the noise furnishes us with the desired jump moment a_n. In practice, this procedure is not terribly edifying, although you may wish to try it as an exercise. You will need to expand the force $f(x[t'])$ around $x(t)$, and identify terms of order $\mathrm{d}t$. On performing the average, you should eventually find that

$$\langle[x(t+\mathrm{d}t)-x(t)]\rangle = \frac{1}{\gamma}f(x[t])\mathrm{d}t + O(\mathrm{d}t^2) \tag{5.31}$$

$$\langle[x(t+\mathrm{d}t)-x(t)]^2\rangle = \frac{\Gamma}{\gamma^2}\mathrm{d}t + O(\mathrm{d}t^2) \tag{5.32}$$

$$\langle[x(t+\mathrm{d}t)-x(t)]^3\rangle = O(\mathrm{d}t^2)\ . \tag{5.33}$$

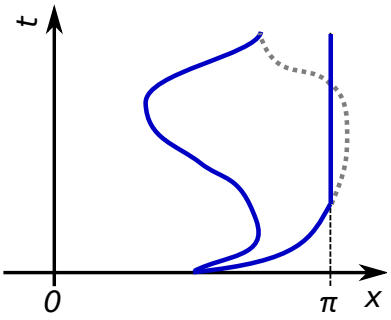


Figure 5.3 *Stochastic motion of a particle in the presence of absorbing boundaries. When the particle hits a boundary point, $x = 0, \pi$, it is absorbed (or sticks there). We can work out the probability that a particle reaches the point (x, t) without hitting a boundary by setting the weight of any path crossing a boundary to zero (i.e., the subpath shown with the dashed line). This is achieved by setting the survival probability to zero at boundary points.*

Therefore we see that the Fokker-Planck equation terminates at the second order term, and indeed has the form (5.29).

5.6 Absorbing boundaries

Another way we can exploit our knowledge of equilibrium dynamics in a nonequilibrium context is to allow the particle to become absorbed (exit the system, or stick) at a boundary, as shown in Figure 5.3. The idea is that until the boundary is reached, the usual equilibrium dynamics are followed. So in the interior region, we still have the same Fokker-Planck equation for the process; but need to modify the dynamics at the boundaries.

It turns out that the correct way to model absorption is to insist that the probability distribution function $P(x, t)$ vanishes at the boundary point(s) at all times $t > 0$. To understand why this is the case, it is best to think of the conditional probability $P(x, t|x_0, t_0)$ as a sum over all paths that go from x_0 to x over a time $t - t_0$. We can divide all paths up into two sets: those that cross the boundary, and those that don't—see Figure 5.3. To get the **survival probability** of the particle that is, the probability that it has reached x, t and *not* crossed the boundary, we want to restrict the sum to the latter set. This happens automatically if the probability to be *at* the boundary is forced to be zero at all times.

One then finds that the total survival probability $P_s(t|x_0, t_0) = \int_{\text{in}} \mathrm{d}x P(x, t|x_0, t_0)$, where the subscript in means 'over the interior region', decreases over time as a consequence of the absorption. It is therefore of interest to calculate this survival probability as a function of time.

5.7 Diffusing particle with two absorbing walls

As a simple example, let's take a diffusing particle that is not subject to an external potential but does stick on contact to walls situated conveniently at $x = 0$ and $x = \pi$, as shown in Figure 5.3. In the interior region $0 < x < \pi$, the particle is governed by the usual Fokker-Planck equation

$$\frac{\partial}{\partial t} P(x, t|x_0, t_0) = D \frac{\partial^2}{\partial x^2} P(x, t|x_0, t_0) \tag{5.34}$$

subject to the initial condition $P(x, t_0|x_0, t_0) = \delta(x - x_0)$. To impose the boundary condition $P(0, t|x_0, t_0) = P(\pi, t|x_0, t_0) = 0$ at all $t > t_0$, we solve the partial differential equation by separation of variables—see Exercise 5B.

Let us suppose we now want to find the total probability that the particle has exited at the right boundary. This is an example of a **first passage property**. One way to obtain this is to integrate the probability flux (4.19)

$$j(x,t) = -D\frac{\partial}{\partial x}P(x,t) \tag{5.35}$$

at the right boundary over a given time interval. As you will discover in the Exercise, this can be done, but it turns out much simpler to calculate such quantities using a different form of the Fokker-Planck equation.

5.8 Backward equation for first-passage properties

In performing the Kramers-Moyal expansion, we stated that instead of making the second time interval infinitesimal, we could have made the first time interval infinitesimal. Following through the derivation leads to the **backward Fokker-Planck equation**

$$\boxed{\begin{gathered}\frac{\partial}{\partial t_0}P(x_1,t_1|x_0,t_0) = -\sum_{n=1}^{\infty}\frac{1}{n!}a_n(x_0,t_0)\frac{\partial^n}{\partial x_0^n}P(x_1,t_1|x_0,t_0) \\ \text{where} \\ a_n(x,t) = \lim_{\delta t\to 0}\frac{\langle[x(t+\mathrm{d}t)-x(t)]^n\rangle}{\mathrm{d}t}\ .\end{gathered}} \tag{5.36}$$

This equation contains the same jump moments as the forward equation (5.26) but there are a number of subtle differences. First, all derivatives are with respect to the *initial* coordinate and time. The terms on the right-hand side all have the same sign. And the jump moments appear outside (rather than under) the derivatives.

What use is such an equation? Let us specialise to the two-wall problem of the previous section. The backward equation reads

$$\frac{\partial}{\partial t_0}P(x,t|x_0,t_0) = -D\frac{\partial^2}{\partial x_0^2}P(x,t|x_0,t_0)\ . \tag{5.37}$$

Now, since the jump moments for this problem are time-independent, the probability distribution depends only on the time difference $t-t_0$, and not the absolute values of these numbers. Hence, derivatives w.r.t. t_0 are equal in magnitude to those w.r.t. t, but take opposite sign:

$$\frac{\partial}{\partial t}P(x,t|x_0,t_0) = D\frac{\partial^2}{\partial x_0^2}P(x,t|x_0,t_0)\ . \tag{5.38}$$

Now, we want the survival probability which is obtained by integrating over the final coordinate x. Now, this integral commutes with the derivatives w.r.t. t and x_0. Hence,

$$\frac{\partial}{\partial t}P_s(t|x_0,t_0) = \frac{\partial}{\partial t}\int_0^\pi \mathrm{d}xP(x,t|x_0,t_0) = D\frac{\partial^2}{\partial x_0^2}\int_0^\pi \mathrm{d}xP(x,t|x_0,t_0) = D\frac{\partial^2}{\partial x_0^2}P_s(t|x_0,t_0)\ . \tag{5.39}$$

That is, the survival probability satisfies the backward Fokker-Planck equation. In fact, so does the exit probability $Q(t|x_0,t_0)$; or indeed any linear transformation of the distribution at the later time t. The quantity of interest implies a particular choice of boundary conditions on the backward equation. For example, the survival probability up to time t must be zero if the initial position is on either boundary. So the appropriate boundary condition on (5.39) is $P_s(t|0,t_0) = P_s(t|\pi,t_0) = 0$. See Exercises 5B and 5C for applications of the backward equation.

5.9* More general stochastic processes

We conclude this lecture with a coda on more general—potentially fully nonequilibrium—processes. The whole point of the Kramers-Moyal expansion is that it allows one to write down the Fokker-Planck equation for any Markov process in continuous space. All we need to do is determine the jump moments. We can also handle discrete processes if we use a continuum limit as an approximation.

For example, let us consider a simple model of population dynamics in which n individuals (mushrooms, say) may reproduce with probability $r\delta t$ in a timestep of length δt (making r a reproduction rate), or die with probability δt (so that the death rate is unity). In the limit $\delta t \to 0$, the population size changes by at most one individual, so $\delta n = \pm 1$, and we find the jump moments are

$$a_1(n,t) = a_3(n,t) = a_5(n,t) = \cdots = n(r-1) \tag{5.40}$$

$$a_2(n,t) = a_4(n,t) = a_6(n,t) = \cdots = n(r+1)\ . \tag{5.41}$$

Notice that in this case, all jump moments are finite, so the corresponding Fokker-Planck equation would have an infinite number of terms. We can approximate the process by first treating the population size as a continuous variable $n \approx x$; and by truncating the Fokker-Planck equation at the second term:

$$\frac{\partial}{\partial t}P(x,t) \approx -(r-1)\frac{\partial}{\partial x}xP(x,t) + \frac{r+1}{2}\frac{\partial^2}{\partial x^2}xP(x,t)\ . \tag{5.42}$$

Notice that this equation has an interesting feature: the diffusion "constant" has picked up a spatial dependence: it increases linearly with x. Since the diffusion constant is proportional to the square root of the magnitude of the fluctuations, we see that the size of the fluctuations, relative to the distance from the origin, decreases as $1/\sqrt{x}$. This is exactly what you would expect: the effect of the fluctuations is proportionately larger in small populations. Another example of a model with a spatially-varying diffusion term is the subject of Exercise 5C.

5.10* Multiplicative noise and the Ito-Stratonovich dilemma

In setting up a problem like the population dynamics model just described, one might be tempted to start with a Langevin description, and build the scaling of the fluctuations with the population size into the noise correlator. That is, to write down to mean behaviour and employ the 'just add noise' method, taking into account that the noise strength should be proportional to $\sqrt{x}$:

$$\dot{x}(t) = (r-1)x(t) + \sqrt{x(t)}\xi(t) \tag{5.43}$$

where the noise satisfies the usual Gaussian properties:

$$\langle \xi(t) \rangle = 0 \quad \text{and} \quad \langle \xi(t)\xi(t') \rangle = \Gamma\delta(t-t')\ . \tag{5.44}$$

There are two problems with this approach. Because we don't have detailed balance underlying these dynamics (the macroscopic population of mushrooms does not satisfy Hamilton's equation), we don't have a fluctuation-dissipation relation to constrain Γ. More seriously, this Langevin equation (5.43), does not correspond uniquely to a single Fokker-Planck equation, but instead to a family of Fokker-Planck equations.

Technically, what happens is the following. Consider integrating a Langevin equation of the form

$$\dot{x}(t) = f[x(t),t] + g[x(t),t]\xi(t) \tag{5.45}$$

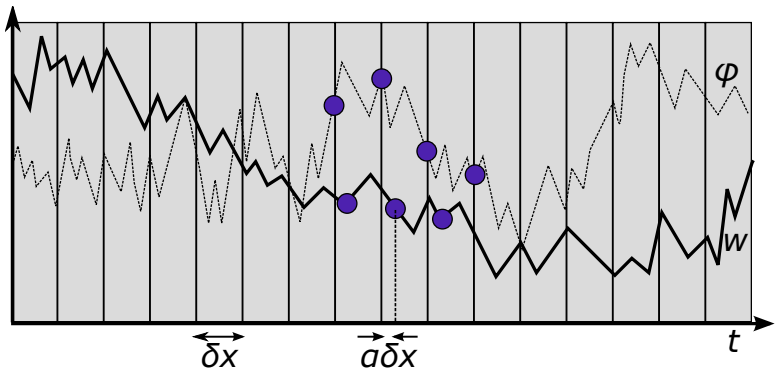


Figure 5.4 *Discretisation of the integral (5.47). Dotted line shows the function $w(t)$, the heavy line the function $g[x(t),t)]$. The values of w are read off at the end-points, and of g a fraction α of the distance between them.*

where $\xi(t)$ has the usual properties. First we find

$$x(t) - x(0) = \int_0^t \mathrm{d}t' f[x(t'),t'] + \int_0^t \mathrm{d}t' g[x(t'),t']\xi(t') \ . \tag{5.46}$$

Now, defining $w(t) = \int_0^t \mathrm{d}t'\xi(t')$, we can write the second integral as the limit of a discrete sum

$$\int_0^t \mathrm{d}t'[g[x(t'),t']\xi(t') = \lim_{n\to\infty} \sum_{i=0}^{n-1} g[(1-\alpha)t_i + \alpha t_{i+1}][w(t_{i+1}) - w(t_i)] \ . \tag{5.47}$$

Now, if the function $w(t)$ were smooth, the choice of α in this definition would not appear in the answer. But because the derivative of $w(t)$ is uncorrelated noise, this is not necessarily true. Specifically, if $g(x,t)$ is constant, the answer is α-independent; whereas if it depends on the random functions x or η it usually is. One can verify this explicitly for the case $g(x,t) = w(t)$, for example.

The question then arises: what is the right choice for α? This is the essence of the **Ito-Stratonovich** dilemma, these being the names given to the two particular choices $\alpha = 0$ and $\alpha = \frac{1}{2}$, respectively. Other choices are possible, and recent experimental work on colloids diffusing near a boundary that affects the diffusion constant suggests that $\alpha = 1$ is the appropriate choice for this problem (Volpe et al., 2010), a fact that has generated some discussion in the literature.

If you ever find yourself encountering **multiplicative noise**, that is, a Langevin equation where the noise term is multiplied by the random function $x(t)$, you should pause, and consult a good textbook (like Risken, 1989) and read up on this dilemma before proceeding. Alternatively, simply start from a process you believe in (for example, a discrete process), calculate the jump moments directly, and avoid Langevin equations—and hence the dilemma—completely.

Exercises

5A: Linear response theory

The mobility of a damped particle is defined as the linear response of its velocity to an applied field. For example, for a particle with charge q in an electric field E with terminal velocity v,

$$\mu = \left.\frac{\partial v}{\partial E}\right|_{E=0}$$

(a) Using the Langevin equation for a damped particle

$$m\ddot{x} + \gamma\dot{x} = qE + \xi(t)$$

derive the **Einstein relation**

$$\mu = \frac{qD}{kT} .$$

(b) Relate the mobility to the equilibrium velocity autocorrelation function $C_{vv}(t) = \langle v(t)v(0)\rangle_0 - \langle v(0)\rangle_0\langle v(t)\rangle_0$ by deriving the **Green-Kubo** relation

$$D = \int_0^\infty \mathrm{d}t C_{vv}(t) . \tag{5.48}$$

[Hint: begin by considering a finite range of integration, and use the fact that at equilibrium $\langle v(t)\rangle_0 = 0$, $\langle v(t)v(t+\tau)\rangle_0 = \langle v(t)v(t-\tau)\rangle_0$ and $\langle v(0)v(t)\rangle_0 = \langle v(\tau)v(t+\tau)\rangle_0$.]

5B: First-passage to a boundary

Consider the situation described in the lecture where a single particle is diffusing between absorbing barriers at $x = 0$ and $x = \pi$.

(a) By separation of variables, show that the general solution to the diffusion equation satisfying $P(0,t) = P(\pi,t) = 0$ is

$$P(x,t) = \sum_{n>0} a_n \sin(nx)\mathrm{e}^{-n^2Dt} .$$

For the initial condition $P(x,t_0) = \delta(x - x_0)$ show that

$$P(x,t|x_0,t_0) = \frac{2}{\pi}\sum_{n>0} \sin(nx_0)\sin(nx)\mathrm{e}^{-n^2Dt} .$$

(b) Calculate the probability current

$$j(x,t) = -D\frac{\partial}{\partial x}P(x,t|x_0,t_0)$$

at the right boundary. Define the probability that the particle has exited at the right boundary by time t as

$$Q(t|x_0,t_0) = \int_{t_0}^t \mathrm{d}t' j(\pi,t') .$$

Obtain its limiting form as $t \to \infty$.

(c) Show that $Q(t|x_0,t_0)$ satisfies the backward Fokker-Planck equation

$$\frac{\partial}{\partial t}Q(t|x_0,t_0) = D\frac{\partial^2}{\partial x_0^2}Q(t|x_0,t_0) . \tag{5.49}$$

[Hint: begin with an expression for the left-hand side in terms of $P(x,t|x_0,t)$, differentiate with respect to time t and at the end integrate back up]. What are the boundary conditions on $Q(t|x_0,t_0)$ at $x_0 = 0$ and π? Use the fact that in the limit $t \to \infty$ the time derivative vanishes to solve this equation, and compare with your answer for part (b).

5C: The Moran model with mutation

The Moran model is a simple model in population genetics (or ecology, where it is called the Hubbell model) in which there are N asexually reproducing individuals of two types (alleles of a gene, or species). In one timestep the dynamics are as follows: a randomly-chosen individual dies, and is replaced by an individual of type 1 with probability m_1; type 2 with probability m_2; or a copy of another individual present in the population with probability $1 - m_1 - m_2$.

(a) Let x be the fraction of individuals of type one in the population at the beginning of a timestep. Calculate the mean and mean-square change in x over one timestep, i.e., $\langle \delta x \rangle$ and $\langle [\delta x]^2 \rangle$.

(b) Show that if m_1 and m_2 remain constant in the limit $N \to \infty$, one obtains a deterministic equation of motion for x (in a continuous time variable where one unit of time corresponds to N timesteps); whereas if m_1 and m_2 scale as $1/N$ in the limit $N \to \infty$, i.e.,

$$m_i = \frac{\mu_i}{N}$$

one obtains a Fokker-Planck equation in a continuous time variable where one unit of time corresponds to N^2 timesteps that terminates at the second term.

(c) Show that the steady-state solution of the equation is

$$P_{\rm ss}(x) \propto x^{\mu_1 - 1}(1-x)^{\mu_2 - 1} \ .$$

Comment on the limiting form as μ_1 and $\mu_2 \to 0$.

(d) For a general Fokker-Planck equation with first and second jump moments $a_1(x)$ and $a_2(x)$ and absorbing boundaries at $x = 0$ and 1, show that the mean time $T(x_0)$ to reach *either* boundary starting from x_0 is the solution of

$$-1 = a_1(x_0)T'(x_0) + \frac{1}{2}a_2(x_0)T''(x_0)$$

subject to the boundary conditions $T(0) = T(1) = 0$. [Hint: write down an expression for T in terms of the probability $P_s(x_0, t)$ neither boundary has been reached by time t and consider the integral $\int_0^\infty \mathrm{d}t\, t\ddot{P}_s(x_0, t)$]

(e) Hence show that, in the absence of mutation $m_1 = m_2 = 0$, the time taken for an allele (or species) with initial frequency x_0 to take over the whole population or go extinct in the Moran model is

$$T(x_0) = -\left[x_0 \ln x_0 + (1 - x_0)\ln(1 - x_0)\right] \ .$$

Lecture 6

Far-from-equilibrium fluctuations

Although most of the foundations of statistical mechanics were established in the late 19th century and have endured to this day, the past twenty years has seen a cluster of exact results relating to fluctuations in systems driven far from equilibrium. These have received a lot of press (see e.g. Bustamante et al., 2005), in part because they are surprisingly easy to derive, because they are broadly applicable, because they are experimentally verifiable and finally because they challenged some widely-help misconceptions about the relationship between microscopic dynamics and macroscopic thermodynamics.

6.1 Ensemble switching and the minimum work principle

We will be dealing with changes to a system, such as its volume or the external field, that are not necessarily reversible. For concreteness, consider a fluid in contact with a heat bath at temperature T, and whose volume V can be altered by moving a piston. The equilibrium state of the system is described by the Helmholtz free energy $F(T,V)$. We can do work on the fluid by moving the piston so that its volume changes from V_0 to V. This process is sometimes called **switching** from the ensemble with volume V_0 to that with volume V.

The laws of thermodynamics imply that the work done on compressing the fluid must satisfy the inequality

$$\Delta W \geq \Delta F = F(T,V) - F(T,V_0) \ . \tag{6.1}$$

Likewise, if work the fluid expands and does work, that work cannot exceed the free energy difference. In either case, the equality holds only if the process is done reversibly; a strict inequality implies that the entropy of the 'universe' (which minimally comprises only the fluid and the heat bath) has increased.

The proof is straightforward. The bath can change its energy only by exchanging heat with the fluid. If an energy ΔW is fed into the combined system (fluid + bath) from outside, and the fluid changes its internal energy by ΔU, it follows the bath has received heat $\Delta Q = \Delta W - \Delta U$ by conservation of energy. The bath is assumed never to change its temperature (this is the defining property of a bath), and hence the entropy change of the combined system is

$$\Delta S = \Delta S_{\rm bath} + \Delta S_{\rm fluid} = \frac{\Delta W - \Delta U}{T} + \Delta S_{\rm fluid} = \frac{\Delta W - \Delta F}{T} \tag{6.2}$$

since $F = U - TS$. The second law of thermodynamics ($\Delta S \geq 0$, with the equality holding only for reversible processes) implies the minimum work principle (6.1).

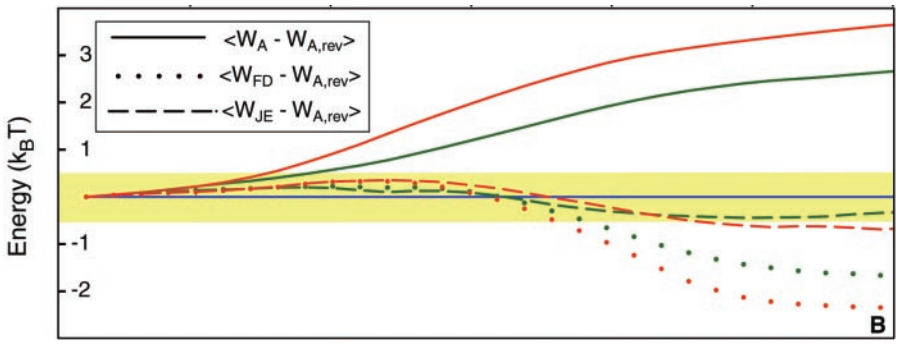


Figure 6.1 *Three estimates of a free energy change of RNA under stretching. W_{A} is the actual work done; W_{FD} is an estimate using the fluctuation-dissipation relation; and W_{JE} an estimate using Jarzynski's equality. The actual free energy difference, as measured in a reversible process, has been subtracted off. Despite dissipation of up to 3kT, the Jarzynksi equality gives results consistent with the equilibrium free energy difference. The horizontal axis is the RNA extension, from* 0 *to* 30*nm (each tick is* 5*nm). Figure from Liphardt et al. (2002).*

For nearly 150 years, the inequality (6.1) was the only general result for irreversible processes. A little over ten years ago, the *equality*

$$\langle \mathrm{e}^{-\Delta W/kT} \rangle = \mathrm{e}^{-\Delta F/kT} \tag{6.3}$$

was published in Physical Review Letters by Jarzynski (1997). It is known, imaginatively enough, as the Jarzynski equality.

The angle brackets denote an ensemble of repetitions of a macroscopic switch between two ensembles (here, two different volumes of a fluid V_0 and V). Usually, it is envisaged that the same macroscopic protocol is followed each time (i.e., the volume $V(t)$ a time t into the switch is the same for each repetition), although the proof requires only that the system always starts at equilibrium with the heat bath at temperature T and volume V_0 and ends with volume V. At the end of the switch, the fluid and the bath are *not* required to be at equilibrium, which is a remarkable feature of the equality (along with the fact that it is an equality). In particular, it allows one to obtain equilibrium free energy differences from *nonequilibrium* measurements.

Figure 6.1, published in Science by Liphardt et al. (2002), shows three different estimates of a free energy change in a strand of RNA that was stretched (worked) by being grafted to polystyrene beads that were manipulated by laser traps. One is simply the work done in the stretching; if done slowly, this should equal ΔF, but done quickly it should exceed it. The upper plot, showing the difference between the work done in the irreversible and reversible experiments, show that this is indeed the case. The fluctuation-dissipation relation can be used to obtain a more refined estimate. This does a better job of estimating the free energy difference, but undershoots at a stretch of about 18nm. The estimate obtained from the Jarzynski equality shows good agreement with the reversible stretch data across the whole range 0–30nm.

The relationship between the Jarzynski equality and the minimum work principle is an interesting one. Jensen's inequality for convex functions implies that $\langle \mathrm{e}^{-X} \rangle \geq \mathrm{e}^{-\langle X \rangle}$ for any random variable X. Hence, from (6.3),

$$\langle \Delta W \rangle \geq \Delta F \,. \tag{6.4}$$

The Jarzynski equality would seem to imply the classical minimum work principle, but only on average. That is, it does not preclude the violation of (6.1) in some instances of the switching protocol. If this were to have occurred, then the entropy of the universe must have decreased in those experiments. The

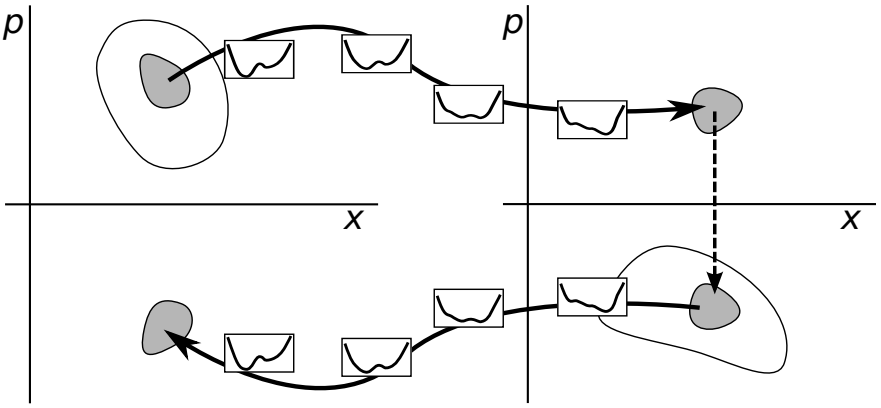


Figure 6.2 *Transformation of phase space volume when work is being done, indicated schematically by the insets which show the evolution of the external potential along a trajectory. The original condition is recovered if we reverse the external work protocol after time-reversing the final configuration. As in Lecture 1, the volume of phase space in the bath is conserved by the dynamics; the detailed balance relation is obtained by expressing the phase space volume corresponding to a particular trajectory in the system of interest with that corresponding to its initial point.*

existence of these 'second law violations' had already been recognised four years previously by Evans et al. (1993), and has formed the basis of much research (and misinterpretation) since.

6.2 Detailed balance relation for an ensemble switch

The key to understanding these recent results is to work out what is going on in terms of heat and work at the microscopic level. The consequence of doing work on a system of interest is that its internal energy function $U(\mathbf{x})$ changes over time. If this system is coupled to a bath, as in Lecture 1, then we can associate changes of energy due to the time-dependence of $U(\mathbf{x})$ with externally applied work; and the remainder to heat exchange with the bath. These ideas are explored in more detail in Exercise 6A.

When the internal energy function was time-independent, we obtained the detailed balance relation between the probability of witnessing a trajectory X and its time reversal $\hat{X}$ from the same initial equilibrium condition. It is possible to extend this result to the case where the initial and final equilibrium conditions are different because the internal energy function has changed due to the applied work. The key to obtaining the detailed balance relation was to realise that the initial condition of the system can be recovered by running the dynamics backwards in time from the final condition. The same is still true when the internal energy is time-dependent: we can recover the initial configuration by running the work protocol backwards in time. For example, if we increase the volume from V_0 to V linearly in the forward process, we obtain the reverse process by decreasing from V to V_0 at the same rate. See Figure 6.2.

As in Lecture 1, we denote $\Omega(X)$ as the number of configurations of the bath at the initial time that lead to the trajectory X being followed under the forward work protocol. We introduce $\hat{\Omega}(\hat{X})$ as the number of configurations at the final time that lead to the reverse trajectory $\hat{X}$ being followed under the reverse work protocol. We have just said these two numbers are equal:

$$\Omega(X) = \hat{\Omega}(\hat{X}) \; . \tag{6.5}$$

As in Lecture 1, we can use this equality to make statements about conditional probability distributions. We are going to compare two conditional probabilities: (i) $P(X|X_0)$, the probability of following trajectory X given initial condition X_0 and equilibrium with the bath at temperature T and the form of the

internal energy function that applies at time $t = 0$; (ii) $\hat{P}(\hat{X}|\hat{X}_0)$, the probability of following the reverse trajectory $\hat{X}$ given it's initial condition $\hat{X}_0$ (which is the time-reversal of the final point reached under the forward dynamics) and starting at equilibrium with the bath at temperature T and the form of the internal energy function that applies at the *final* time $t > 0$. Using the fact that all configurations are equally likely in the respective equilibrium conditions, we again find that

$$\frac{P(X|X_0)}{\hat{P}(\hat{X}|\hat{X}_0)} = \frac{\hat{\Omega}[\hat{X}_0]}{\Omega[X_0]} \tag{6.6}$$

where $\Omega[X_0]$ is the number of bath configurations compatible with the initial condition X_0 at time $t = 0$, and $\hat{\Omega}[\hat{X}_0]$ the corresponding quantity at the later time t. Again, Boltzmann's formula implies that

$$\boxed{\frac{P(X|X_0)}{\hat{P}(\hat{X}|\hat{X}_0)} = \exp\left(\frac{\Delta Q}{kT}\right)}\,, \tag{6.7}$$

where ΔQ is the heat surrendered by the system of interest to the bath in the forward process.

Notice that the derivation of detailed balance here is exactly the same as it was in Lecture 1. The only difference is that during the reverse process, the work protocol is the reverse of that used in the forward process (rather than there being no work protocol).

6.3 Crooks' relation

The extended detailed balance relation (6.7) quickly leads to Crooks' relation (Crooks, 2000). Let us consider the subset of all trajectories X that correspond to an amount of work ΔW done on the system of interest. Starting from equilibrium we can calculate the probability of doing this amount of work:

$$P(\Delta W) = \sum_X P_{\rm eq}(X_0) P(X|X_0)\,, \tag{6.8}$$

where the sum is restricted to the subset of interest. Now, the equilibrium distribution of the initial coordinate X_0 is

$$P_{\rm eq}(X_0) = \exp\left(-\frac{U(X_0) - F}{kT}\right) \tag{6.9}$$

where F is the (Helmholtz) free energy of the equilibrium system at the initial time t. Using (6.7) we find that

$$P(\Delta W) = \sum_X \exp\left(-\frac{U(X_0) - F - \Delta Q}{kT}\right) \hat{P}(\hat{X}|\hat{X}_0)\,. \tag{6.10}$$

Now, the internal energy at the end of the process is $U(\hat{X}_0)$, which from the first law we know to be

$$U(\hat{X}_0) = U(X_0) - \Delta Q + \Delta W\,, \tag{6.11}$$

where we have used the convention that ΔQ is heat absorbed by the bath, and ΔW is the work done on the system of interest. Hence we obtain

$$P(\Delta W) = \sum_X \exp\left(-\frac{U(\hat{X}_0) - F - \Delta W}{kT}\right) \hat{P}(\hat{X}|\hat{X}_0) \tag{6.12}$$

$$= \sum_X \hat{P}_{\rm eq}(\hat{X}_0) \hat{P}(\hat{X}|\hat{X}_0) \exp\left(-\frac{\Delta F - \Delta W}{kT}\right)\,, \tag{6.13}$$

where we have used the form of the equilibrium distribution that applies at the endpoint

$$\hat{P}_{\mathrm{eq}}(X_0) = \exp\left(-\frac{U(\hat{X}_0) - \hat{F}}{kT}\right) \tag{6.14}$$

in which $\hat{F}$ is the equilibrium free energy of the final state; and $\Delta F = \hat{F} - F$.

Finally we note that the exponential term in (6.13) does not depend on the summation variable X and can be taken outside the sum; and that the sum itself is the total probability of following a trajectory $\hat{X}$ under the reverse dynamics. Since under the forward dynamics, each trajectory X is assumed to involve an amount ΔW of work done on the system, under the reverse dynamics each trajectory $\hat{X}$ implies an amount of work $-\Delta W$ done on the system. Hence we obtain **Crooks' relation**

$$\boxed{\frac{P(\Delta W)}{\hat{P}(-\Delta W)} = \exp\left(\frac{\Delta W - \Delta F}{kT}\right)} \tag{6.15}$$

where $\hat{P}(\Delta W)$ is the probability that an amount ΔW of work is done on the system under the reverse dynamics. This relation quantifies the rarity of "second-law violations", that is, microscopic trajectories in which heat is spontaneously transformed into work. It is due to this rarity that they are never observed macroscopically.

6.4 From Crooks' relation to the Jarzynski equality

The Jarzynski equality now follows quickly from (6.15). We rearrange it is

$$P(\Delta W)\mathrm{e}^{-\Delta W/kT} = \hat{P}(-\Delta W)\mathrm{e}^{-\Delta F/kT}\,. \tag{6.16}$$

Now, we integrate over all ΔW. Hey presto:

$$\langle \mathrm{e}^{-\Delta W/kT} \rangle = \mathrm{e}^{-\Delta F/kT}\,. \tag{6.17}$$

There are other ways to derive the Jarzynski equality that don't start from detailed balance, but use the fact that at any instant, the dynamics of the system of interest induced by the heat bath want to restore equilibrium. One such derivation is covered in Exercise 6B.

It is worth reiterating that although the equilibrium distribution at the final state enters into the *derivation* of Crooks' relation, and hence the Jarzynski equality, when using the equality in practice the system does not need to be at equilibrium at the end of the switching protocol. As we have mentioned before, this provides a method by which equilibrium free energy differences can be measured by nonequilibrium experiments. However, the method is only practical if the free energy difference is small, i.e., a few kT. The reason for this is that a large, negative ΔW contributes a large exponential factor to the average in (6.3). Even if the probability that the system spontaneously does work against the driving force is exponentially small in $-\Delta W$, it is still possible that when weighted by the exponential factor $\mathrm{e}^{-\Delta W/kT}$, one obtains an important $O(1)$ contribution to $\mathrm{e}^{-\Delta F/kT}$. Indeed, this turns out to be true in practice, and hence the number of repetitions of the switching process that need to be performed to get a good estimate of ΔF increases exponentially with the typical amount of work done by the switch.

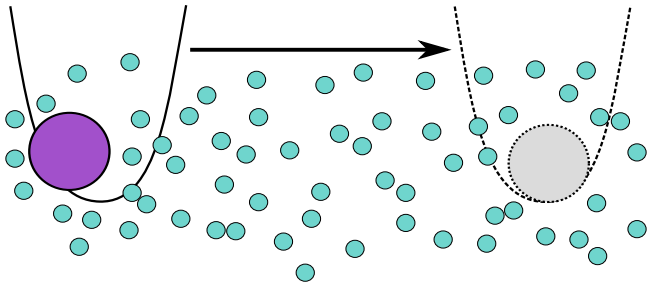


Figure 6.3 *Experimental setup to test Crooks' relation. A colloidal particle is dragged through a fluid by a laser trap which can be modelled as a linearly-translating potential.*

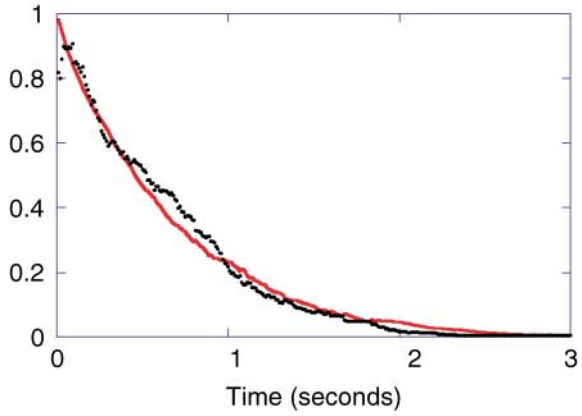


Figure 6.4 *Points: left-hand side of (6.19). Line: right-hand side of (6.19). Data obtained for a trapped colloid dragged through a solvent. Figure from Wang et al. (2002).*

6.5 Direct experimental test of Crooks' relation

We finally remark that Crooks' relation (6.15) has itself been experimentally tested. The experiment in question has a colloidal particle suspended in a fluid and dragged by a laser trap—in other words, a potential is dragged from one place to another as shown in Figure 6.3. Here, the change in the free energy is zero, since the potential has only been moved, but not changed in functional form. In this scenario, the forward and reverse protocols are related by a spatial reflection. So if dragging from left to right is the same thing as dragging from right to left (i.e., the background fluid is isotropic), then the forward and reverse protocols are identical, and we find from (6.15) for this experiment that

$$\frac{P(\Delta W)}{P(-\Delta W)} = \mathrm{e}^{\Delta W/kT} \,. \tag{6.18}$$

It turns out that to get good statistics, it is better to work with an integrated version of this relation:

$$\frac{P(\Delta W < 0)}{P(\Delta W > 0)} = \langle \mathrm{e}^{-\Delta W/kT} \rangle_+ \,. \tag{6.19}$$

Here, $\langle \cdot \rangle_+$ means an average over only the positive work trajectories. Figure 6.4 is taken from Wang et al. (2002) who performed precisely the experiment just described. In this figure, the left- and right-hand sides are plotted on the same axes and show excellent agreement, thereby confirming the validity of Crooks' relation.

Exercises

6A: Microscopic heat and work exchanges during an ensemble switch

Consider an overdamped Brownian particle in a time-dependent harmonic potential

$$U(x,t) = \frac{\kappa(t)}{2}x^2 \ , \qquad \kappa(t) = \begin{cases} \kappa_1 & 0 < t < t_s \\ \kappa_2 & t_s < t < t' \end{cases}$$

arising from a switching protocol running from time $t = 0$ to $t = t'$.

(a) Write down the work done by the switch as a functional of the particle trajectory $x(t)$, likewise the change in internal energy and hence the heat surrendered to the bath.

(b) Write down conditions under which the entropy of the bath decreases as a consequence of the switch.

(c) Assume that the particle is at equilibrium with the bath at time $t = 0$, and take $t' \to \infty$ so that it ultimately equilibrates with the new potential. By equipartition, or otherwise, show $\langle \Delta U \rangle = 0$. Further show that

$$\langle \Delta W \rangle = \langle \Delta Q \rangle = \frac{kT}{2}\left(\frac{\kappa_2}{\kappa_1} - 1\right) \ .$$

(d) Show that

$$\Delta F = \frac{kT}{2}\ln\frac{\kappa_2}{\kappa_1}$$

and hence that the minimum work principle is satisfied.

(e) Explicitly verify the Jarzynski equality $\langle \mathrm{e}^{-\Delta W/kT} \rangle = \langle \mathrm{e}^{-\Delta F/kT} \rangle$ for this switching protocol.

6B: Jarzynski equality in the presence of local heating

In the lecture we derived the Jarzynski equality for a system in contact with a heat bath via Crooks' relation. A more direct proof, which highlights the fact that both system of interest and the bath may go out of equilibrium during the ensemble switch, is due to Pradhan et al. (2008) and is the subject of this exercise.

Let $\mathbf{x}$ and $\mathbf{y}$ denote the coordinates of particles in the system of interest and bath respectively. The momenta of the particles will not enter into any calculations so we shall ignore them. The Hamiltonian of the combined system we will thus write as

$$H(\mathbf{x},\mathbf{y}) = H_\mathrm{s}(\mathbf{x},t) + H_\mathrm{b}(\mathbf{y}) + H_\mathrm{int}(\mathbf{x},\mathbf{y}) \ .$$

Here, only any external work is allowed only to affect the degrees of freedom in the system of interest ('s'), hence why the Hamiltonians for the bath ('b'), or system-bath interaction ('int') do not exhibit time-dependence.

(a) Consider a time interval Δt during which $H_\mathrm{s}(\mathbf{x},t)$ is time-independent. We shall label this interval n, and define the set transition probabilities $P_{\Delta t}(\mathbf{x},\mathbf{y} \to \mathbf{x}',\mathbf{y}')$, where $\mathbf{x},\mathbf{y}$ is the state of the system at the start of the interval and $\mathbf{x}',\mathbf{y}'$ that at the end. Show that they must satisfy

$$\int \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} P_{\Delta t}(\mathbf{x},\mathbf{y} \to \mathbf{x}',\mathbf{y}')P_n^*(\mathbf{x},\mathbf{y}) = P_n^*(\mathbf{x}',\mathbf{y}')$$

where $P_n^*(\mathbf{x},\mathbf{y})$ is the stationary (equilibrium) distribution w.r.t. the Hamiltonian that applies during the time interval n:

$$P_n^*(\mathbf{x},\mathbf{y}) = \frac{1}{\mathcal{Z}_n}\mathrm{e}^{-H(\mathbf{x},\mathbf{y})/kT} \ .$$

(b) If we allow a work 'event' to occur at time t_n, defined to occur at the end of interval n, show that the work done is

$$\Delta W_n = H_s(\mathbf{x}[t_n], t_n^+) - H_s(\mathbf{x}[t_n], t_n^-)$$

where $t_n^\pm$ are intended to indicate times just before and after the work event. Hence show that,

$$\int \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}\mathrm{e}^{-\Delta W_n/kT} P_{\Delta t}(\mathbf{x},\mathbf{y}\to\mathbf{x}',\mathbf{y}') P_n^*(\mathbf{x},\mathbf{y}) = \frac{\mathcal{Z}_{n+1}}{\mathcal{Z}_n} P_{n+1}^*(\mathbf{x}',\mathbf{y}') \ .$$

(c) By chaining together a sequence of N such intervals, and integrating over an equilibrium distribution at time $t = 0$, show that

$$\langle \mathrm{e}^{-\Delta W/kT} \rangle = \frac{\mathcal{Z}_N}{\mathcal{Z}_1} \ .$$

Show that the Jarzynski equality then follows if the free energy of the combined system is the sum of that for the two components, system of interest and bath, separately. Explain why this derivation holds if the system of interest and/or bath go out of equilibrium during the switching process.

6C: Quasi-static limit of a continuous switching protocol

We consider the same harmonic potential set-up as in Exercise 6A, but now allow $\kappa(t)$ to be a continuously-varying function of time.

Show that if initially the distribution of the particle's position is Gaussian with zero mean and known variance, it remains Gaussian for all times with zero mean and variance $w(t) = \langle x(t)^2 \rangle$ that satisfies the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t} w(t) = \frac{2}{\gamma}[kT - \kappa(t) w(t)] \ .$$

[Hint: consider the time evolution of the Fourier transform of $P(x,t)$ w.r.t. x].

Show that the mean work done is

$$\langle \Delta W \rangle = \frac{1}{2} \int_0^t \mathrm{d}t' \dot{\kappa}(t') w(t') \ .$$

Taking a linear profile for $\kappa(t)$ between two values κ_0, κ_1 over a duration t, calculate the mean work done, and demonstrate explicitly the convergence to the quasi-static limit as $t \to \infty$.

TOP

Topological Phases

Physics by the Lake, August 2018

Dr Sam T. Carr, S.T.Carr@kent.ac.uk

Contents

1 Introduction: the quantum Hall effect **1**

2 Band theory in two dimensions and the Chern number **3**
2.1 Chern index / number for two dimensional band models 4
2.2 Impossibility of global gauge and practical calculation of the Chern number 5
2.3 The quantum Anomalous Hall effect 7
2.3.1 Two band system 7
2.3.2 The Haldane model 9
2.4 Edge states 13
2.4.1 Chiral modes 15
2.4.2 The Bulk-Boundary correspondence 17

3 Symmetry protected topological states and the Su-Schreiffer-Heeger model **17**
3.1 A mathematical description of the winding number 20
3.2 Edge states 20
3.3 The protecting symmetry 21

4 Time reversal symmetry and the quantum spin Hall effect **23**
4.1 The Kane-Mele model 23
4.2 Helical edge states 23
4.3 The role of time reversal symmetry 25
4.3.1 Kramers pairs 26
4.3.2 $\mathbf{Z}_2$ topology 26

5 The periodic table of topological insulators and superconductors **27**
5.1 Non-unitary symmetries 27
5.2 Classification 29

6 Beyond topological band theory **30**

7 Further reading **31**
7.1 Lecture notes available on the web 31
7.2 Books 31
7.3 Review articles 32

8 Workshop questions **32**

Preface

The study of topological phases of matter have been rising in popularity over the last few years, and this popularity will only increase with the award of the 2016 Nobel prize to this field. These lecture notes aim to introduce the reader to topological band theory and symmetry protected topological phases. There are many good books or reviews on the topic, and the point of these lecture notes is to start at the beginning and guide the reader through the first steps so that they can then pick up a more comprehensive textbook to get deeper into the subject. As well as getting an overview of the field, we will also go into detail of some calculations which are often skipped in the textbooks.

1 Introduction: the quantum Hall effect

Let us begin with recalling a few things about the (integer) quantum-Hall effect, which is a two-dimensional electron gas in a magnetic field. The quantum-Hall effect is a huge subfield of condensed matter physics on its own, but for the purposes of introducing topological states of matter, I only want to think about the edge states.[1] There are two elementary ways to think of the edge modes, which are illustrated in Fig. 1

- From the semi-classical point of view, the magnetic field makes the electrons undergo circular motion, known as cyclotron orbits. This is all very well, except for the edge of the sample, where the electron doesn't have space to complete the orbit, and instead bounces off the edge. The electron then propagates around the edge in the opposite direction to the cyclotron orbit, known as a skipping orbit. This is one way to visualise the chiral edge modes present in quantum-Hall insulators.

- An alternative viewpoint is fully quantum, thinking about Landau levels. In the bulk, the energy bands are completely flat (in the absence of disorder), known as Landau levels. However, at the edge of the sample, these states have to go somewhere – in the standard textbook picture, they all go off to $E \to +\infty$, indicating that there is a confining potential $V \to +\infty$. Therefore, wherever you put the chemical potential (the filling fraction in the quantum-Hall effect), the Landau levels cross the chemical potential at the edge of the sample, giving rise to metallic edge modes.

The first point of view is a very nice visualisation of the physics of the edge mode in a magnetic field, but ultimately can't explain why there is a quantised number of skipping orbits going all the way around the sample. For example, if the sample edge wasn't completely straight and got wider at some point, what is to stop the orbit skipping along until it gets to the wider bit to continue with cyclotron motion at that point? While undergraduate textbooks come up with lots of nice (and at least partially correct) ways of explaining this, ultimately the question is always going to be there as to why this edge conductance is so well quantised – something that one answer in the semi-classical picture.

Likewise, the second point of view is a very nice (and also more or less accurate as far as it goes) way of viewing what happens to Landau levels when a confining potential is added – but if one blindly accepts this point of view, one will (or at least, should!) ask the question: Why doesn't this happen for any insulating state? Why don't we see edge modes in most band insulators?

[1] I hope you've seen these concepts before in an undergraduate course, but even if not, then the gist is easy to grasp.

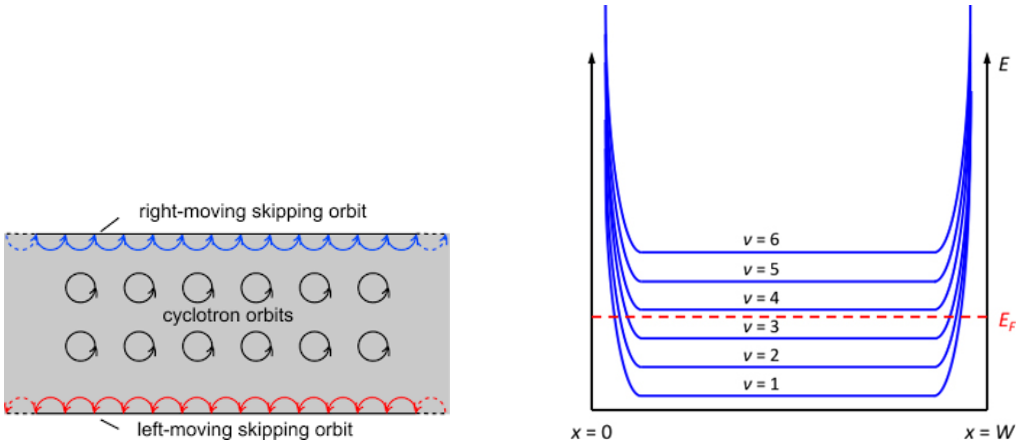


Figure 1: Left: Semi-classical view of edge states in the quantum-Hall effect as skipping orbits. Right: Quantum view of edge states as bending of Landau levels up to infinity at the edge of the sample – at either edge, 3 Landau levels cross the Fermi level, giving three metallic modes at the edge, even through the bulk is an insulator.

Ultimately, this is related to topology, which was realised in a famous paper by Thouless, Kohomoto, Nightingale and den Nijs.[2] They calculated a quantity for each Landau level labelled by m, which is the integral over all momentum of the Berry curvature $\mathcal{F}_m$ of the wavefunction in that Landau level,

$$n_m = \frac{1}{2\pi}\int d^2k \mathcal{F}_m \tag{1}$$

They then defined a quantity N which is the sum of n_m over all occupied Landau levels. So long as the chemical potential is in a gap between Landau levels, so each one is either completely occupied or not, it turns out N is an integer, which is both the number of edge states, and also the Hall conductance $\sigma_{xy} = \frac{e^2}{h}N$. This quantity N is a topological number and answers both the questions we posed above – in particular, it explains why the edge states are such perfect conductors that they are now used as the experimental metrological definition of the quantum of conductance. We will see how all of this fits together as we go through the course.

It turns out however that while it was the first topological state known and is still by far the most important, it is not the easiest to understand. This is because there are two length scales in the problem – the unit cell,[3] and the magnetic length, which is in general incommensurate with the unit cell. These extra complications turn out not to be important for understanding the topology of the model, however they do add extra complications, so I will actually not discuss the quantum-Hall effect in any detail in these introductory lectures, but will rather start with an almost equivalent effect known as the anomalous quantum Hall effect, which doesn't require the presence of an external magnetic field.

The field of topological band theory in more general terms than the quantum (anomalous) Hall effect however really started later around 2005, with the discovery of the quantum spin Hall effect, which is an example of a symmetry protected topological

[2] Physical Review Letters, 49, 405 (1982)

[3] The presence of a periodic lattice turns out to be important for topological physics – we will see why later

phase. The aim of these lectures is to introduce you first to the anomalous quantum Hall effect and introduce a little bit about the theory of topology, and then go on to look at these more modern aspects of it.

2 Band theory in two dimensions and the Chern number

Consider a Hamiltonian in a periodic potential. Then the wave functions may be written in the Bloch form

$$\psi(k,x) = u(k,x)e^{\mathrm{i}k\cdot x}, \tag{2}$$

Here, k and x are two-dimensional vectors. In the wave function, the Bloch momentum k is an index labelling the state in question. In the case of a periodic potential in continuous space, the Bloch wave function $u(k,x)$ is a periodic function $u(k,x) = u(k,x+a)$ where a is one of the lattice vectors. Alternatively for a lattice model, $u(k,x)$ is not a function of x, but is rather a vector, with length given by n the number of lattice sites in the unit cell. All examples in this course will be on a lattice, but the generalisation to continuous wave functions is rather trivial.

After Fourier transformation, the Hamiltonian is a square matrix $H(k)$ of size $n \times n$. For each Bloch momentum therefore, we have an eigenvalue equation

$$H(k)u_m(k) = \epsilon_m(k)u_m(k). \tag{3}$$

For each value of k, there will be n different solutions of the equation labelled by $m = 1 \ldots n$; these are the bands, and for each band m and value of k, there is the energy $\epsilon_m(k)$ and the eigenstate $u_m(k)$.

- Pre 1980s – diagonalize the Hamiltonian to get the band structure energies $\epsilon_m(k)$. There was never a reason to get the associated wavefunctions; except perhaps to calculate matrix elements or expectation values of various operators. All thermodynamics comes from the energies alone.
- Post 1980s (or there about) – it was realised that the states (wave functions) give additional information about the system that characterise distinct phases of matter, for example the quantum Hall phase.

These phases aren't characterised by spontaneous broken symmetry and a local order parameter as in the Landau theory. Indeed, we are currently dealing with non-interacting models so phase transitions of this form are not possible. Rather, the distinct phases are characterised by a topological index, which comes from the wave functions.

We will investigate this by looking at the simplest such topological index; the Chern index (or Chern number, sometimes called the first Chern number). This is not the only possibility; in the final lecture, we will look at another one.

2.1 Chern index / number for two dimensional band models

We consider the states in a Bloch band $\underline{u}_m(\underline{k})$, as defined above. To avoid overcomplicating the expressions, we will not always write the band index; however one must remember that each of these things is defined independently for each band, as they are calculated from the wave functions which naturally depend on the band in question.

The **Berry connection** is defined as

$$A_\nu(\underline{k}) = i\langle u(\underline{k})|\partial_{k_\nu}|u(\underline{k})\rangle \tag{4}$$

where $\nu = x, y$. If we have a lattice model and our states are simply vectors, then this expression becomes

$$A_\nu(\underline{k}) = i\underline{u}^* \cdot \frac{\partial \underline{u}}{\partial k_\nu} \tag{5}$$

If we had a model in continuous space, then the inner product would involve a spatial integral over the unit cell.

The Berry connection is **not gauge invariant**. Consider a gauge transformation on the wave functions:

$$u'(k) = u(k)e^{if(k)}. \tag{6}$$

No observable will depend on this gauge transformation – indeed; all we have done is change the phase we have chosen for each different Bloch wavefunction (i.e. for each different value of Bloch momentum k). However, under such a transformation, the Berry connection changes

$$A'(k) = A(k) - \nabla f(k). \tag{7}$$

In contrast, the **Berry curvature**

$$\mathcal{F}(k) = \frac{\partial A_y}{\partial k_x} - \frac{\partial A_x}{\partial k_y} \tag{8}$$

is gauge invariant – it is easy to see that the above gauge transformation doesn't affect $\mathcal{F}$.

The **Chern index** or **first Chern number** of the band C is then given by integrating the curvature over the entire band

$$C = -\frac{1}{2\pi}\int_{\mathrm{BZ}} d^2k\mathcal{F}(k), \tag{9}$$

where the integral is over the entire Brillouin zone. It is worth here remembering that we started with the wave functions of a given band $u_m(k)$, so the Chern index is something that one associates with an entire band, and depends on the band in question. We will therefore sometimes label the band subscript C_m when it is needed.

2.2 Impossibility of global gauge and practical calculation of the Chern number

Using Green's theorem in a plane, it may appear that one can write

$$C = \int_{\mathrm{BZ}} d^2k\mathcal{F}(k) = \int_{\mathrm{BZ}} d^2k \left(\frac{\partial A_y}{\partial k_x} - \frac{\partial A_x}{\partial k_y}\right) = \oint_{\mathrm{BZ}} dk \cdot A(k), \tag{10}$$

where the final integral is around the boundary of the Brillouin zone. There is an important issue with this – the Bloch momentum is defined on a taurus (i.e. $k = \pi$ is equivalent to $k = -\pi$), which means that the Brillouin zone doesn't really have a boundary. Indeed, if one tries to integrate around a square Brillouin zone (for example), then the integral on one of the sides exactly cancels the integral on the other side, because the two sides actually represent the same state, see Fig. 2. It would seem that the Chern number is always zero.

There is however a rather subtle problem with this statement, which is that it assumes that $A(k)$ is a smooth function of k. It may be however, that there is no global gauge $f(k)$ that one could choose that makes $A(k)$ smooth over the entire Brillouin zone. While this statement may seem rather surprising at first, we will shortly see an explicit example of it. It is also worth noting at this point that this lack of continuity plagues the gauge

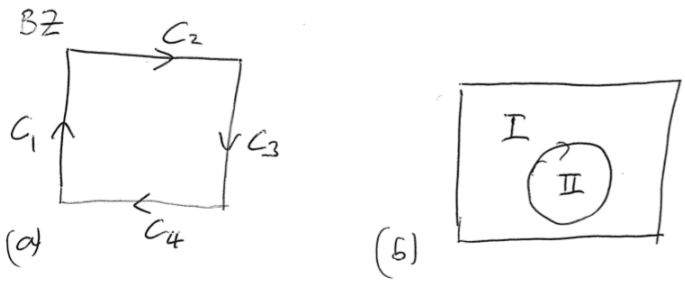


Figure 2: Left panel: an example of a contour around the edge of a Brillouin zone. However, the Brillouin zone is periodic, so the left edge and right edge are actually the same, meaning that the integral along C_1 and C_3 of anything must cancel, as they are the same integral. The same thing happens for C_2 and C_4. Right panel: an illustration of splitting the Brillouin zone into two regions. The integral around the outer edge of the square cancels as in the left panel, so one is left with the integral of the size of the discontinuity in the function around the boundary of region II.

dependent Berry connection $A(k)$, but not the Berry curvature $\mathcal{F}(k)$. Indeed, as $\mathcal{F}(k)$ is gauge independent; then patching gauges will not change $\mathcal{F}$, and the fundamental definition of Chern number (9) always holds.

If one can't choose a global phase so that $A(k)$ is continuous everywhere, then one can patch the system. Take most of the Brillouin zone to be patch I, with Berry connection $A_{\rm I}(k)$. Around the part that has a discontinuity, make a second patch II, with a Berry connection $A_{\rm II}(k)$ given by a gauge transformation $u_{\rm II}(k) = u_{\rm I}(k)e^{{\rm i}\alpha(k)}$ chosen in some way to get rid of the singularity in $A_{\rm I}$; so that

$$A_{\rm II}(k) = A_{\rm I} - \nabla f(k). \tag{11}$$

We can then split the integral in (9) into the two patches, as illustrated in Fig. 2:

$$\int_{\rm BZ} d^2k\mathcal{F}(k) = \int_{\rm I} d^2k\mathcal{F}(k) + \int_{\rm II} d^2k\mathcal{F}(k). \tag{12}$$

Now, because $A_{I(II)}$ is continuous in patch $I(II)$, then we can use Green's theorem in the plane on each of these integrals:

$$\int_{\rm BZ} d^2k\mathcal{F}(k) = \oint_{\rm BZ} dk\cdot A_{\rm I}(k) - \oint_{\rm II} dk\cdot A_{\rm I}(k) + \oint_{\rm II} dk\cdot A_{\rm II}(k). \tag{13}$$

The first of these integrals is zero (as we showed above – the Brillouin zone doesn't really have an edge), while the second integral about our chosen patch depends only on the difference $A_{\rm II}(k) - \cdot A_{\rm I}(k) = -\nabla f(k)$, where $f(k)$ is the gauge transformation we chose in patch II. Hence

$$\int_{\rm BZ} d^2k\mathcal{F}(k) = -\oint_{\rm II} \nabla\alpha\cdot k = -\left[\alpha(\theta=2\pi) - \alpha(\theta=0)\right]. \tag{14}$$

where θ parameterises the angle around the boundary of patch II.

Now, $\theta = 2\pi$ and $\theta = 0$ are the same place, however f is a phase in the wave function, so $\theta(2\pi)$ and $\theta(0)$ only need to be the same modulo 2π for the wave function to be single

valued (as it must be). Hence this integral must be 2π times an integer; which means that the Chern number, (9) is an integer as promised. We also see that it is the winding number of the gauge transformation needed to patch the discontinuous part of the Berry connection somewhere on the Brillouin zone.

It is also worth mentioning that if the discontinuity is due to a bad choice of initial gauge, and the patch wasn't really needed, then the winding number of the gauge transform (assuming that the gauge transform doesn't add any new singularities!) would just be zero – the patching calculation will always work so long as A is continuous in each of the patches.

Some comments:

1. Berry curvature is gauge independent and experimentally observable – this is uniquely defined and always smooth. The patching is only needed to apply Greens theorem in the plane to (9) in order to calculate it in an efficient way (as well as to deeper understand what exactly is the Chern number).

2. The Chern number becomes undefined when there is an extra degeneracy – i.e. when two bands touch, or in other words, when the gap closes. The Chern number is therefore only defined for each band in an insulator. While it may seem strange that changing the eigenstate at a single point can change the Chern number, as we have seen above, a single discontinuity in the Berry connection is all it requires (in fact, the essential requirement!) in order to calculate the Chern number.

3. The Chern number is a map from the wave functions to an *integer*. It is is clear that smoothly changing the wave function must smoothly change the Berry curvature, so must smoothly change the Chern number. However, an integer can not be smoothly changed, which gives us one of the most crucial points:

   The Chern number is invariant under smooth changes of the wave function; so long as the gap to other bands remains non-zero.

   In other words, the Chern number labels distinct insulating states of matter, that may only be deformed into each other by closing the gap.

4. Sometimes the Chern number is directly observable – for example in quantum Hall and quantum anomalous Hall, the Hall conductivity $\sigma_{xy} = (e^2/h)C$ is directly proportional to the Chern number. This is useful for metrology applications, as C is necessarily an integer.

5. There is something known as the *Bulk–boundary* correspondence. We will come back to this later, but roughly speaking, if one imagines two different integers with different Chern numbers placed next to each other, at the interface where the Chern number changes, the gap must disappear (as we have just argued). This means that along the boundary/interface between the two materials, there must be metallic (gapless) modes. This also applies to the edge of the sample, which one can imagine in a similar vein as being a boundary between the sample (with some Chern number C), and the vacuum with Chern number 0).

2.3 The quantum Anomalous Hall effect

As an example of this, we will look at an effect known as the *quantum Anomalous Hall effect*, which was originally discussed by Haldane in 1988 for which he recently won the Nobel prize. The starting point of the model is the tight binding model on a hexagonal

lattice (i.e. graphene, but Haldane's paper was well before Graphene became popular!), where the Hamiltonian is given by

$$H(\underline{k}) = \begin{pmatrix} 0 & 2t_1\cos(\sqrt{3}k_x/2)e^{-ik_y/2} + e^{ik_y} \\ 2t_1\cos(\sqrt{3}k_x/2)e^{ik_y/2} + e^{-ik_y} & 0 \end{pmatrix} \tag{15}$$

This was introduced in the mesoscopic physics lectures, and we recall that the 2×2 structure of the matrix is due to the two sub-lattices of Graphene, and the fact that the matrix is completely off-diagonal is because the hopping is only between A and B sublattices. This 2×2 structure is why Graphene has two bands[4] Haldane's model adds an extra hopping term to graphene which we will look at in detail shortly, however it is worth looking first in a little more detail at the general structure of two-band systems.

2.3.1 Two band system

We recall the Schrödinger equation in momentum representation:

$$H(k)u_m(k) = \epsilon_m(k)u_m(k). \tag{16}$$

For a two band system, $H(k)$ is a 2×2 Hermitian matrix, and we will drop the k index unless it is explicitly required. We all know that a 2×2 Hermitian matrix can be paramaterized by 4 real numbers:

$$H = \begin{pmatrix} a & b + ic \\ b - ic & d \end{pmatrix}. \tag{17}$$

However, there is a more convenient parameterisation in terms of the Pauli matrices [5] :

$$H = h_0\sigma_0 + \underline{h} \cdot \underline{\sigma} = \begin{pmatrix} h_0 + h_z & h_x - ih_y \\ h_x + ih_y & h_0 - h_z \end{pmatrix}. \tag{18}$$

The eigenvalues of the Schrodinger equation with a matrix paramaterised this was are

$$\epsilon_\pm = h_0 \pm |\underline{h}| = h_0 \pm \sqrt{h_x^2 + h_y^2 + h_z^2}. \tag{19}$$

This can either be verified by direct computation (which is a worthwhile exercise if you are out of practice in such things), or by realising that this general Hamiltonian is exactly the Zeeman splitting of a state of energy h_0 in a magnetic field of magnitude $|h|$, as the direction of the field is irrelevant for the energy splitting.

The direction of the field does however change the eigenstates (i.e. wavefunctions) of the Hamiltonian, and this is the important part for classifying topology. There is however one simplifying feature of wavefunctions, which is that they do not depend on the constant shift h_0. We are therefore going to assume $h_0 = 0$ and work with[6]

$$H = \underline{h} \cdot \underline{\sigma} = \begin{pmatrix} h_z & h_x - ih_y \\ h_x + ih_y & -h_z \end{pmatrix}. \tag{20}$$

Let us focus on the lower eigenvalue (lower band) only, so writing it explicitly, we want the associated eigenvector of the eigenvalue $\epsilon_- = -|\underline{h}|$. We will look at two ways to do this.

[4] Of course, graphene itself is a semi-metal rather than an insulator as the bands touch at two points, however the structure of two bands is still there.

[5] Just for completeness, σ_0 is the identity $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, and $\underline{\sigma} = \begin{pmatrix} \sigma_x & \sigma_y & \sigma_z \end{pmatrix}^T$ is the vector of Pauli matrices, $\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\sigma_y = \begin{pmatrix} 0 & -i \\ - & 0 \end{pmatrix}$, and $\sigma_x = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

[6] This in essence is just identifying where is the zero of energy, however I leave it to the reader to think about what might happen if $h(0)$ is a function of k.

1. If you find an appropriate maths textbook, you will find that one of the most common approaches is to write the field in spherical polar coordinates:

$$\underline{h} = |\underline{h}| \begin{pmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{pmatrix}, \tag{21}$$

which gives (you can check this explicitly) the eigenvector:

$$u_- = \begin{pmatrix} \sin(\theta/2)e^{i\phi/2} \\ -\cos(\theta/2)e^{-i\phi/2} \end{pmatrix} \tag{22}$$

There are a couple of comments to make on this:

a) The eigenvector depends only on the direction of $\underline{h}$, and not on its magnitude. This is always true and certainly obvious when you think about it; however it is worth thinking about it explicitly briefly here as it will be far less obvious in other ways of writing the eigenvectors.

b) If one sweeps $\underline{h}$ around in a full circle (e.g. $\theta \to \theta + 2\pi$, or similarly for ϕ), then $u_- \to -u_-$.

This second point is extremely important, as if one insists that the wavefunction is uniquely defined for any $\underline{h}$, then u_- as a function of the direction of $\underline{h}$ will have a discontinuity somewhere. This in turn means that if the Hamiltonian is a smooth function of $\underline{k}$, then u_-, then the wavefunction as a function of k will have a discontinuity somewhere. This is not in fact a problem with this particular parameterisation of the Hamiltonian and the wavefunction, but rather a *feature* of the mapping from 2×2 Hamiltonians to their eigenvectors. This is in fact exactly what we were talking about in Sec. 2.2 about the lack of a global gauge, which we will now develop further.

2. While spherical polar coordinates are in principle very nice to characterise a point on the surface of a sphere (i.e. the direction of $\underline{h}$), they aren't so convenient for practical calculation. A much more practical alternative is to directly use the coordinates of $\underline{h}$, where direct calculation will find that the vector

$$u_-^I = \frac{1}{N} \begin{pmatrix} h_z - |\underline{h}| \\ h_x + ih_y \end{pmatrix} \tag{23}$$

satisfies the Schrödinger equation. Here, N is a normalisation constant, and this is all very well so long as $|\underline{h}$ is not in the positive-z direction, in which case $u_- = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and is not normalisable. Alternatively, one can also check that

$$u_-^{II} = \frac{1}{\tilde{N}} \begin{pmatrix} -h_x + ih_y \\ h_z + |\underline{h}| \end{pmatrix} \tag{24}$$

also satisfies the Schrödinger equation (where $\tilde{N}$ is a different normalisation constant), and this function is fine so long as $|\underline{h}$ is not in the negative-z direction.

While you can (and should) explicitly check that both u_-^I and u_-^{II} satisfy the (same) Schrödinger equation for the lower eigenvalue ϵ_-, this should alarm you, as the normalised eigenvector of a Hermitian matrix is supposed to be unique. Actually, the non-sleep-deprived amongst you will realise that this last sentence wasn't completely precise, and what it really should say is that the normalised

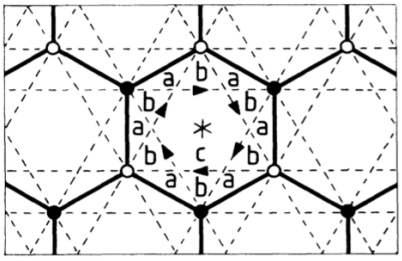


Figure 3: A schematic of Haldane's tight-binding model. The empty and filled dots relate to the A and B sublattices; the a and b in the figure refer to regions with effective positive (negative) flux, which relates the model more directly to the quantum Hall effect but won't be discussed here. Figure taken from Haldane, PRL 1988

eigenvector of a Hermitian matrix is supposed to be unique *up to a phase*. In fact, u_-^I and u_-^{II} differ only by a phase[7]

$$e^{i\alpha} = \frac{(h_z + |\underline{h})/(h_x + ih_y)}{|(h_z + |\underline{h})/(h_x + ih_y)|}. \tag{25}$$

I leave it as an exercise for the reader to (a) check that the expression above is indeed just a phase; and (b) show that $u_-^{II} = e^{i\alpha} u_-^I$.

What this second way of writing the eigenvectors has shown us, aside from practical calculation aspects, is that one can define a smooth mapping from $\underline{h}$ (i.e. the space of all 2×2 Hamiltonians) to the lowest eigenvector everywhere except for one point. This point was the positive-z direction for u_-^I and the negative-z direction for u_-^{II}, but symmetry tells us that this given any point, we could construct a wavefunction (or more precisely, a guage) for which the mapping from the Hamiltonian to the wavefunction is smooth everywhere except this one point.

If we combine this with what we learned about the Chern number is Sec. **??**, this means if $H(k)$ translated into a direction $\underline{h}(k)$ misses just one direction in the Brillouin zone, then the Chern number is zero. If however $\underline{h}(k)$ points in every direction, then one cannot construct a global gauge where the happing from the Hamiltonian $H(k)$ to the wavefunction $u_-(k)$ is smooth everywhere, and we will therefore have a Chern number $C \neq 1$. We will now see an explicit example of this; the Haldane model, which was the first time this was demonstrated in physics beyond the quantum Hall effect, and won Haldane the Nobel prize.

2.3.2 The Haldane model

Haldane's model is a tight binding model of spinless Fermions on a hexagonal lattice, illustrated in Fig. 3 There are three terms in the Hamiltonian:

1. A hopping of strength t_1 between nearest neighbors on the lattice.

2. An energy of $\pm M$ for Fermions sitting on the $A(B)$ sublattices; and

[7] when both are normalisable – and if only one was normalisable then there wasn't a problem anyway!

3. A complex next-nearest neighbor hopping $t_2 e^{\pm i\phi}$, where the positive sign is in the direction of the arrows shown in Fig. 3.

Before going on to solve the model fully, there are a few comments worth making:

1. If $M = t_2 = 0$ (i,.e. only the first term above is present), then this is the tight-binding model for Graphene, which you have some familiarity with. Graphene is a semi-metal, and is not topological.

2. The second term M breaks the sub-lattice symmetry, making A and B sublattices different.

3. The third term breaks time reversal symmetry so long as $\phi \neq 0, \pi$. For any other ϕ, an electron going around a loop in one direction will acquire a different phase than if it goes round the loop in the opposite direction – hence changing the direction of time changes the model. This is similar to a particle in a magnetic field[8] however the phases cancel out on average meaning that it is a bit different from having a magnetic field. The parts of the lattice labelled a and b in Fig. 3 have positive (negative) flux through them, as one would get from a magnetic field in the positive (negative) z direction. The two areas however average out to zero so there is no net magnetic field (and hence no Landau levels).

Let us now construct the Hamiltonian for the Haldane model – as there are two sublattices, there will be a 2×2 structure to the Hamiltonian, which we will write in k space parameterised as $H(\underline{k}) = \underline{h}(\underline{k}) \cdot \underline{\sigma}$. Let us start with the off-diagonal elements, which connect the A sublattice to the B sublattice. The only term that does this is the nearest neighbor hopping, which is identical to graphene. We can therefore go back to our graphene Hamiltonian Eq. (15), where we see this is parameterised as

$$h_x = t_1 \left[2\cos\left(\frac{\sqrt{3}k_x}{2}\right)\cos\left(\frac{k_y}{2}\right) + \cos\left(k_y\right) \right] \tag{26}$$

$$h_y = t_1 \left[-2\cos\left(\frac{\sqrt{3}k_x}{2}\right)\sin\left(\frac{k_y}{2}\right) + \sin\left(k_y\right) \right] \tag{27}$$

The other two terms in the Hamiltonian are diagonal in the sub-lattice basis, and performing the Fourier transform of the positions on the hexagonal lattice (which is left as an exercise) gives us

$$h_z = M - 2t_2 \sin\phi \left[2\sin\left(\frac{\sqrt{3}k_x}{2}\right)\cos\left(\frac{3k_y}{2}\right) - \sin\left(\sqrt{3}k_x\right) \right]. \tag{28}$$

This describes the Hamiltonian of the Haldane model, where $\underline{k}$ runs over the Brillouin zone of the hexagonal lattice.

Now we have the model, we need to analyse it. The energy spectrum may be easily obtained from the general expression Eq. (19)

$$\epsilon_\pm(k) = \pm|\underline{h}|. \tag{29}$$

This can be plotted for various values of the parameters (and I encourage you to do so!), but nothing is gained from writing the long general expression. Leaving it in this

[8] indeed, one adds a magnetic field to a lattice model via the Peierls substitution, which adds a phase $e^{i\int_A^B \underline{A}\cdot\underline{r}}$ to the hopping.

form, we see that if the chemical potential is set at $\mu = 0$ (which is the charge neutrality point in the context of graphene), then we have a two band insulator, unless the gap closes somewhere, which can only happen if $|h(k)| = 0$ somewhere, i.e. $h_x(k) = h_y(k) = h_z(k) = 0$ for some k.

Let us start with pure graphene, i.e. $t_2 = M = 0$ which means that $h_z = 0$. Then a little algebra shows us that $h_x = h_y = 0$ at

$$k = \pm \begin{pmatrix} 4\pi/3\sqrt{3} \\ 0 \end{pmatrix}, \tag{30}$$

which are known as the K and K' points, or sometimes simply just the Dirac points, of graphene.[9] If one assumes we are close to one of these points in momentum space,

$$k = \pm \begin{pmatrix} 4\pi/3\sqrt{3} \\ 0 \end{pmatrix} + \begin{pmatrix} q_x \\ q_y \end{pmatrix}, \tag{31}$$

one can expand the cosines and sines for small $\underline{q}$ to get

$$h_x + ih_y = \frac{3t_1}{2} \left(\pm q_x + iq_y \right), \tag{32}$$

in other words we have the Dirac Hamiltonian $H = v\underline{\sigma} \cdot \underline{q}$, where $v = 3t_1/2$ is the Fermi-velocity (or speed of light for the Dirac Hamiltonian).[10]

Let us now assume $t_2, M \neq 0$, so we go away from graphene. Let us continue to focus on the region around the Dirac points, i.e. expand for small $|q|$:

$$h_z(K + q) = M + 3\sqrt{3}t_2 \sin\phi \tag{33}$$

$$h_z(K - q) = M - 3\sqrt{3}t_2 \sin\phi, \tag{34}$$

In other words, the low energy expansion near the two Dirac points of graphene give

$$H_\pm = v\underline{\sigma} \cdot \underline{q} + m_\pm \sigma_z, \tag{35}$$

where the mass term $m_\pm = M \pm 3\sqrt{3}t_2 \sin\phi$.

Unless one of $m_\pm = 0$, in which case there is no gap around one of the K points, the system is therefore a fully gapped insulator. However, there are still different cases which can be classified by the Chern number, as we now see. First let us assume $M < 0$. Then there are two cases:

1. $|M| > 3\sqrt{3}t_2 \sin\phi$.

   In this case, $h_z < 0$ at both the K and K' points. This means that for the entire Brillouin zone, there is no value of $\underline{k}$ where $\underline{h}$ is in the positive-z direction.[11] This means we can use the wavefunction u^I_- from Eq. (23) everywhere, and hence the Chern number must be 0.

[9] You will of course find other solutions to $h_x = h_y = 0$, but you will also find that they are equal to either the K or K' point plus some reciprocal lattice vectors; i.e. they are the same point in the Brillouin zone.

[10] To be precise, it is only the Dirac Hamiltonian at the K point, while at the K' point, it should be $H = v(-\sigma_x q_x + \sigma_y q_y)$. One can turn it into the Dirac Hamiltonian by making the substitution $q_x \to -q_x$ at the K' point. As the calculations demonstrated in these notes are all done around the K point, one doesn't need to worry about this; however if you repeat the calculation at the K' point, this extra $-$ sign should be taken into account, as it does make a difference in some quantities.

[11] This finally illustrates why we don't miss anything by the small q expansions – for $\underline{h}$ to point in the z direction, we must have $h_x = h_y = 0$, which can only happen at the K points, so all we need to do is look near there.

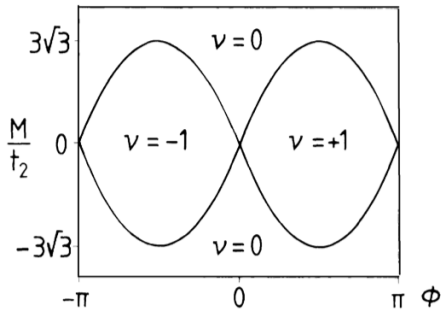


Figure 4: The phase diagram of Haldane's model. Figure taken from Haldane, PRL 1988. Haldane's original paper denoted the Chern number as η, in analogy with the filling fraction in the Quantum Hall Effect, while we have been calling it C.

2. $|M| < 3\sqrt{3}t_2 \sin\phi$.

   In this case, $h_z < 0$ at the K' points, but $h_z > 0$ at the K point. This means we can't use u^I_- everywhere, as it is not normalisable at the K point – however we can use this wavefunction everywhere except a small circle around the K point, where we will need to use u^{II}_-. The gauge transformation in this region can be obtained by substituting the Hamiltonian, Eqs. (32) and (34) into the general formula we derived earlier, Eq. (25), which gives

$$e^{i\alpha} = \frac{m_+/v(q_x + iq_y)}{|m_+/v(q_x + iq_y|} = \frac{|q_x + iq_y|}{q_x + iq_y} = e^{-i\theta}, \tag{36}$$

   where θ is the polar angle of the momentum, $q_x + iq_y = |q|e^{i\theta}$. Finally, putting this into Eq. (14) to calculate the integral of the Berry curvature over the Brillouin zone gives us

$$\int_{\mathrm{BZ}} d^2k\mathcal{F}(k) = -\oint_{\mathrm{II}} \nabla\alpha \cdot k = -\left[\alpha(\theta = 2\pi) - \alpha(\theta = 0)\right] = 2\pi, \tag{37}$$

   which tells us that the Chern number,

$$C = \frac{1}{2\pi}\int_{\mathrm{BZ}} d^2k\mathcal{F}(k) = 1. \tag{38}$$

One can repeat the calculation for other regions of parameter space, which gives the phase diagram shown in Fig. 4 One sees three phases as a function of M and ϕ, distinguished by different Chern numbers $C = -1, 0, 1$.

It is worth making a few remarks on this result:

1. This was the first topological insulator discovered beyond the quantum Hall effect; thus showing the topology of the ground state is more general than just this one effect, something that wasn't really appreciated until the topological insulator revolution almost twenty years later. In modern terminology, this is known as the anomalous Quantum Hall effect – there is no external magnetic field, however time reversal symmetry is broken explicitly in the model (something we will return to shortly).

2. More general models like this one where the Chern numbers $C \neq 0$ are known as Chern insulators. While this particular model does not have any phases with $|C| > 1$, it is not difficult to construct models with Chern number equal to any integer you want (indeed, the correct plateaux of the quantum Hall effect would be one example of this).

3. While Haldane's paper was entirely theoretically motivated, the quantum anomalous Hall effect has been realised recently in experiments, albeit in sometihng described by a somewhat more complicated model. For further details of this, see Liu, Zhang and Qi, arXiv:1508.07106.

2.4 Edge states

So far, we have taken a model of an insulator, and shown that one can define a topological quantity, the Chern number, for the ground state of the system. We have seen that there are various phases, separated by lines where the model is metallic (i.e. there is no band gap), which can be characterised by different Chern numbers. Though this is a nice mathematical exercise, you would be perfectly justified in now asking "so what?"

To be more precise, we saw that the Haldane model is similar to Graphene, but with each Dirac point gapped, with gaps $m_\pm$. We have seen that the Chern number depends on the relative signs of the gaps, and hence one can imagine two phases where the magnitude of the gaps $|m_+|, |m_-|$ are the same, but the Chern number is different. What physical properties of the system depend on the Chern number?

1. If one looks at thermodynamic quantities in an infinite system,[12] then there is no difference at all – the different phases are thermodynamically equivalent, as the energy spectrum is the same in each case.[13]

2. If however one looks at thermodynamics in a finite system with edges, one sees extra gapless modes (which will turn out to be localised at the edges) in the case where $C \neq 0$ which aren't present if $C = 0$.

The above comparison is to make a point, that the energy spectrum of the model actually depends on whether the system has edges or not – and in fact, if there are edges, then the state is not exactly insulating as there are gapless excitations localised at the edge. We already talked about this briefly in the introduction, where we looked at skipping orbits in the quantum Hall effect – so maybe we aren't too surprised. However, it is worth also thinking about this from the point of view of numerical work, where one often chooses boundary conditions (periodic or otherwise) in order to minimise finite size effects when trying to extrapolate to the thermodynamic limit. The presence of edge states means that by changing boundary conditions, you might actually also change the physics involved.[14]

Let us first understand edge states in Chern insulators at a very general level. First of all, imagine an interface between two systems with opposite signs of, say m_-. One

[12] or alternatively, a finite system with periodic boundary conditions, as is common in numerical work

[13] While the two states may have the same energy spectrum in the bulk, the wavefunctions are different. This means that correlations functions, or anything that depends on the wavefunctions rather than just the energies may show the difference between the phases. The best example in Chern insulators (including the quantum Hall effect and Haldane model) is the Hall conductance σ_{xy}. From a Kubo-formula calculation which is in all of the textbooks, it turns out that the Hall conductance is in fact a direct measure of the Chern number $\sigma_{xy} = \frac{e^2}{h} C$.

[14] Conversely, this is an excellent way numerically to look for topological states, where you look at how low-lying eigenvalues change when the boundary conditions are changed.

somehow tunes external parameters in the model so that on one side of the sample (say $y < 0$), $m_- > 0$, on the other side ($y > 0$) $m_- < 0$, as illustrated in Fig. **??** One would then say at $y = 0$, we must have $m_- = 0$, hence along the interface between the two phases, we have a gapless mode.

While this line of reasoning gives the basic idea,[15] it is not completely precise. Our original statement involves a function $m_-(y)$ which goes from being positive to being negative. It can only do this by passing through zero right? Unfortunately, not right. There is no good reason that m_- must be real[16] – we may have just chosen a gauge in which it is real. However, something else happens between the two phases, the Chern number also changes.[17] We have already seen that the Chern number is an integer, and is continuously related to the wavefunction through the integral of the Berry curvature (9), except when two wavefunctions become degenerate, i.e. the model becomes gapless. Therefore, the fact that the integer Chern number changes as a function of y means that there must be some value of y where C changes discontinuously, i.e. the system is gapless. This is the *topologically protected edge mode.*

Let me make three comments here:

1. While I have so far talked about edge modes in an interface between two insulators with different Chern numbers, one can also talk about the interface between a single Chern insulator and the vacuum, which has Chern number $C = 0$. Hence if the insulator has $C \neq 0$ and one has a finite piece of it, then there will be topologically protected edge modes at the edge (well, where else would edge modes be?) This explains one of the mysteries we set ourselves in the introduction – in Fig.(to be inserted)

2. The 'topologically protected' part is important, although colloquially we often just say 'edge mode'. The point is that edge modes can appear for a number of reasons, particularly in simple models. The topologically protected part however means that no matter how you perturb the model, so long as you don't close the gap, the edge mode must be there. In contrast, an edge state that arises for non-topological reasons will typically be localised by disorder.[18] We will return to this point shortly.

3. The role of topology in physics has become quite vogue over recent years, and can be used in many completely different contexts. However, one of the basic definitions of (differential) topology is a continuous mapping from continuous functions onto a discrete set of possibilities. In this case, the continuous functions are the wavefunctions; the discrete set is the integers (hence a common term, $\mathbb{Z}$ topological insulator – we will come back to this); and the continuous mapping is the integral of the Berry curvature – the Chern number. You can compare this to commonly drawn pictures showing the number of holes a three dimensional object has – a sphere has none, a taurus has one, etc... A shape is continuous, the number of holes is discrete, and making local perturbations to the shape (without poking any

[15] And to be honest, I actually think it is a good way of thinking of it, so long as one understands the caveat we are about to mention.

[16] Remember that the actual gap is $|m_-|$ – as soon as we allow it to be negative, there is no reason why it can't also be complex in principle.

[17] In fact, to be precise, the fact that the Chern number changes means that in fact, the mass in this model can be constrained to be real – but this is because of the topological quantum number and one can construct other models in which this is not the case. For example, if one sets $\phi = 0$ and crosses the $M = 0$ axis, then both m_+ and m_- change sign, but there doesn't have to be edge states in this case.

[18] Meaning that it might exist in a clean system, but not a disordered one

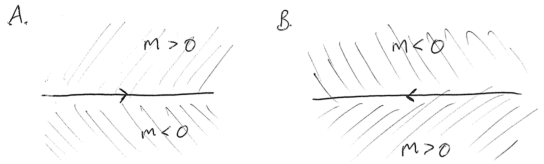


Figure 5: Two illustrations of boundaries in real space between a phase with $m > 0$ and a phase with $m < 0$.

extra holes through the surface which is clearly not continuous) can not change the number of holes. Basically speaking, the idea behind topological protection of anything is that an integer cannot change continuously.

2.4.1 Chiral modes

The previous discussion has illustrated why there must be edge states in a system with Chern number $C \neq 0$. Let us now look in a little more detail at the nature of these edge modes.

Rather than consider the full Haldane model, let us just focus on one of the Dirac cones, which can be described by a Hamiltonian (35), making the usual substitution back $\underline{k} \to -i\nabla$:

$$H = -i\sigma_x \frac{\partial}{\partial x} - i\sigma_y \frac{\partial}{\partial y} + m(y)\sigma_z. \tag{39}$$

We have allowed for a boundary in the y direction, where $m(y)$ changes sign as we cross the line $y = 0$. It is easy to verify two eigenstates of this Hamiltonian $H\psi = E\psi$:

$$\psi_1(x,y) = e^{ik_x x} e^{-\int_0^y m(y')dy'} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad E = k_x \tag{40}$$

$$\psi_2(x,y) = e^{ik_x x} e^{+\int_0^y m(y')dy'} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \qquad E = -k_x \tag{41}$$

It should also be clear that the first wavefunction ψ_1 is only normalisable if $m(y > 0) > 0$ and $m(y < 0) < 0$ as illustrated in Fig. 5A, while the second wavefunction ψ_2 is only normalisable if $m(y > 0) < 0$ and $m(y < 0) > 0$ as illustrated in Fig. 5B. Finally, one can see from the dispersion relations that the velocity of the excitation in the first case $v_{\text{edge}} = \frac{\partial E}{\partial k_x} > 0$, while in the second case $v_{\text{edge}} < 0$. In other words, the edge modes are *chiral*. It doesn't take too much thought to see that if the boundary isn't just defined straight at $y = 0$ but some more complex shape, the edge mode will just go along the boundary in one direction.

While we already met chiral edge modes when we discussed skipping orbits in the introduction, it is worth thinking briefly about them from the point of view of band theory. Suppose we wanted to make a low energy effective one-dimensional model of the edge, for energies far below that of the bulk gap of the material. If we write a normal energy spectrum defined in a Brillouin zone, with $E(k + 2\pi) = E(k)$, then it must cross the chemical potential an even number of times.[19] However, the chiral edge states crosses it just once, as shown in Fig. 6.

A chiral mode is particularly robust, as one can see from the following viewpoints:

[19] Or alternatively, just touch the chemical potential, which is a Lifshitz transition and irrelevant to the present discussion.

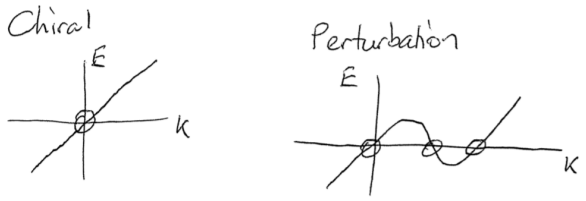


Figure 6: A schematic illustration of the dispersion of a chiral mode. If one makes a continuous deformation of the spectrum, then one must add an even number of crossings of the Fermi level.

1. If one adds disorder to the edge, then naively one expects there to be some resistance, as an impurity can backscatter the electron. In this case however, there is no state to backscatter into – all of the electrons go in the same direction.[20] Hence edge conduction is perfect, $G = e^2/h$, which has been measured to very high accuracy in the quantum Hall effect, and is in fact used to define the quantum resistance standard in metrology.

2. Now imagine that one adds a perturbation to the system (e.g. again from disorder), which changes the energy spectrum of the edge states – see Fig. 6. Then extra states may cross the Fermi energy – but they can only appear in pairs. You may now think that there is a state with negative velocity to scatter into – however, one can now think of the converse and imagine that localisation somehow gaps the system. Again, the Fermi points can only be taken away in pairs, so there is always one left.

In fact, let us define $N_\pm$ as the number of Fermi points with $\pm$ velocity. Then the chirality of the state,

$$Q = N_+ - N_- \tag{42}$$

can not change under any continuous deformation of the energy spectrum, as new Fermi points can only be added in pairs: one with positive velocity and one with negative velocity. Hence Q is a topological quantum number under the definition we wrote down above – it maps something continuous (the energy spectrum of an edge) onto an integer number, and cannot be changed under any continuous deformation.

2.4.2 The Bulk-Boundary correspondence

We have now defined two different topological quantum numbers we can associate with the Haldane model – the Chern number C, which is a measure of the bulk properties of the wavefunctions; and Q which is to do with chiral modes at a boundary. In fact, these two quantum numbers are related. At a boundary between two distinct insulating states, the chirality of the edge mode is given by

$$Q = \Delta C, \tag{43}$$

[20] Strictly speaking, there is a state going the other way, at the other side of the sample. Matrix elements between the two edges of the sample however will be exponentially small in sample size.

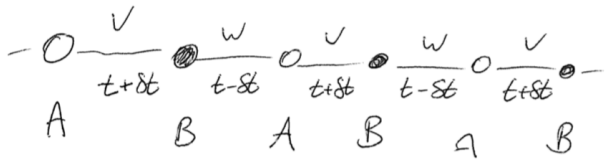


Figure 7: A schematic of the SSH model. The model is a one-dimensional tight-binding chain, with a staggered hopping. The unit cell therefore contains two atoms, which we call label sublattice A and B.

where $\Delta C = C_1 - C_2$ is the change in Chern number across the boundary. You can find a general proof of this in any of the books or lecture notes cited.[21] This is something known as the Bulk-Boundary correspondence, because it relates something about the bulk (which can be calculated in an infinite system) to something that must happen at a boundary. This is a generic feature of topological insulators, and in this particular case, tells us that if we have any system with a Chern number that is not zero[22], there must be $|C|$ chiral edge states, where the sign of C gives the chirality.

3 Symmetry protected topological states and the Su-Schreiffer-Heeger model

So far, we have looked at one topological index, the Chern number, which is relevant for a two-dimensional system. We now turn to another topological index, the winding number. This is rather simpler to visualise, and it also introduces another extremely important concept – that of symmetry protected topological states. To do this, we will look at a model due to Su, Schreiffer and Heeger,[23] which was originally motivated by a Peierls instability in the conducting polymer, polyacetylene. While this is very interesting and I encourage you to read about it, it is irrelevant for the point at hand, so we will just jump straight into the model.

The Hamiltonian is described by a one-dimensional tight binding chain with a staggered hopping, which may be written in second quantised notation

$$H = \sum_n (t + (-1)^n \delta t) \left[c_n^\dagger c_{n+1} + c_{n+1}^\dagger c_n \right], \tag{44}$$

and is illustrated schematically in Fig. 7. The unit cell consists of two atoms, so it is more convenient to write the Hamiltonian in terms of unit cells with two sublattices:

$$H = \sum_n \left[v (c_n^A)^\dagger c_n^B + w (c_n^B)^\dagger c_{n+1}^A + h.c. \right], \tag{45}$$

where we have written the intra-cell hopping as $v = t + \delta t$, and the inter-cell hopping as $w = t - \delta t$. By taking a Fourier transform to write this in momentum space, we end up

[21] This typically uses a construct known as a Thouless Charge Pump, which is very interesting, but takes us too far from the main purposes of the lectures so I won't go into it here.

[22] Because zero is the Chern number of the vacuum

[23] Though not the original reference, the best place to read about some of the early work on this is the review paper Heeger, Kivelson, Schrieffer and Su, Rev. Mod. Phys. 60, 781 (1988).

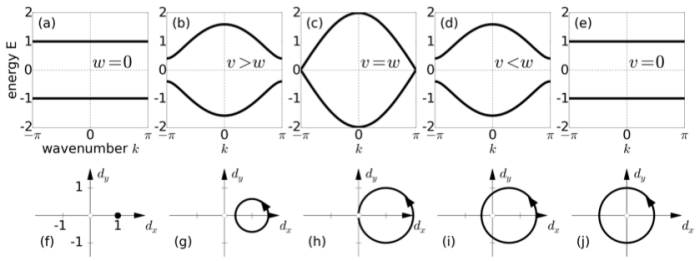


Figure 8: Upper panel – the spectrum of the SSH model as u and v are varied – notice e.g. that cases (b) and (d) are thermodynamically equivalent. Lower panel – tracing the vector $\underline{d}(k)$ that describes the Hamiltonian as a function of k.

with a 2×2 matrix, as we were dealing with in the previous section, which can in turn be written in terms of Pauli matrices:

$$H(k) = \begin{pmatrix} 0 & v + we^{-ik} \\ v + we^{ik} & 0 \end{pmatrix} = \underline{d}((k) \cdot \underline{\sigma}, \tag{46}$$

where

$$d_x = v + w\cos k, \quad d_y = w \sin k, \quad d_z = 0. \tag{47}$$

This of course gives us our spectrum,

$$E(k) = |\underline{d}(k)| = \sqrt{v^2 + w^2 + 2vw\cos k}, \tag{48}$$

which is plotted in Fig. 8 for various values of v and w. From the figure (and the symmetry in the spectrum), it is clear that interchanging the hoppings v and w gives identical spectra, and hence thermodynamic equivalence.

However, let us now also look at the wavefunctions, which as in the case of the Haldane model, we will parameterise by the direction $\underline{d}(k)$ of the Hamiltonian itself. As $d_z = 0$ always, the Hamiltonian (and by extension, the wavefunction, as described in Sec.2.3.1) can be represented by a point on the plane. One can therefore make a parametric plot of $\underline{d}$ as a function of k, which is also shown in Fig. 8 for different values of v and w. Based on these figures, let us attempt to construct a topological quantum number step by step:

1. On the $\underline{d}$ diagram, the origin is a singular point. This is because if $\underline{d} = 0$, then the two energy bands are degenerate, and the wavefunctions can not be defined uniquely.

2. The plot of $\underline{d}(k)$ must be a closed loop, as $k = -\pi = k = \pi$. In other words, it is a map from the Brillouin zone, which is a circle, to the plane.

3. In fact, we could go further and recall that only the direction of $\underline{d}$ matters for the wavefunction, not the magnitude (which gives the energy), so it is actually a map from a circle to a circle. This is another way of saying the first point – that we don't want to pass through the origin.

4. The obvious topological index is therefore the winding number – the number of times $\underline{d}(k)$ wraps around the origin. This is clearly a topological index as it (a) is an integer; and (b) is insensitive to any smooth deformation of $\underline{d}$, so long as this deformation does not go through the origin.

We can see in the diagram that if $w < v$, (i.e. the intra-cell hopping dominates), then the topological index is zero. While if $w > v$ (i.e. the inter-cell hopping dominates), then the index is 1. The transition point $u = v$ is where the spectrum becomes gapless and the system is not an insulator. The topological index is defined differently in this case, however you see the same properties in it as we already saw in the Haldane model.

There is one big difference though. The reason we could represent $\underline{d}$ on a two-dimensional plane is because $d_z = 0$ in the model. But a topological index is supposed to be robust to any perturbation – so what happens if we perturb the model slightly, to have $d_z \neq 0$. In this case, the mapping is not from a circle (the Brillouin zone) to a circle (the space of gapped two-band Hamiltonians with $d_z = 0$), but from a circle to a sphere (the space of gapped two-band Hamiltonians in general). The winding number is no longer a topological index, as one can deform a circle around the origin to a circle one side of the origin without going through the origin by coming out of the plane $d_z = 0$. This is very distinct from the Chern number[24] which put no restrictions on the Hamiltonian.[25]

The SSH model is only topological therefore so long as h_z is forced to be always 0. Now the way to force a Hamiltonian to have fewer free parameters than it might is by imposing a symmetry on the system – and that is indeed the case here. We will come back shortly to look at exactly what the protecting symmetry is that forces $h_z = 0$, but let us take this as our first example (and in many ways, a definition) of a *symmetry protected topological state*. What this means is that there is a topological index which cannot be changed as long as the system remains gapped, *so long as one doesn't break the protecting symmetry*. This allows for a far richer variety of behaviours beyond the quantum Hall effect and quantum anomalous Hall effect, which don't require any symmetry. In fact, outside two dimensions, the only topological states are symmetry protected ones – but we are getting ahead of ourselves and will return to this at the end.

3.1 A mathematical description of the winding number

This is not really important to follow the gist of these notes, but for completeness, let me write down mathematically how to compute the winding number, which we have only discussed so far schematically. We are already assuming that $d_z = 0$, so let us write[26]

$$h(k) = d_x(k) - id_y(k) \tag{49}$$

Then the winding number W is given by

$$W = \frac{1}{2\pi i}\int_{-\pi}^{\pi} dk \frac{h'(k)}{h(k)} = \frac{1}{2\pi i}\int_{-\pi}^{\pi} dk \, \frac{d}{dk}\log h(k) = \frac{1}{2\pi}\left[\theta(\pi) - \theta(-\pi)\right], \tag{50}$$

where in the last step, we have defined $h = |h|e^{i\theta}$ in polar coordinates. The last step follows because $h(\pi) = h(-\pi)$ so the real parts cancel, and hence W is real and an integer. If you want a fun exercise in contour integration, substitute the SSH model into this integral and verify that W is what we expect it to be.

[24] Where the difference is that we were mapping from a taurus to a sphere, rather than a circle to a sphere

[25] In fact, one can prove that there is no possibility of defining any topological index of a mapping from a circle to a sphere, but more on that later.

[26] An alternative way of defining this is to simply write the Hamiltonian as $H(k) = \begin{pmatrix} 0 & h(k) \\ h^*(k) & 0 \end{pmatrix}$, which explains the minus convention in the formula in the main text.

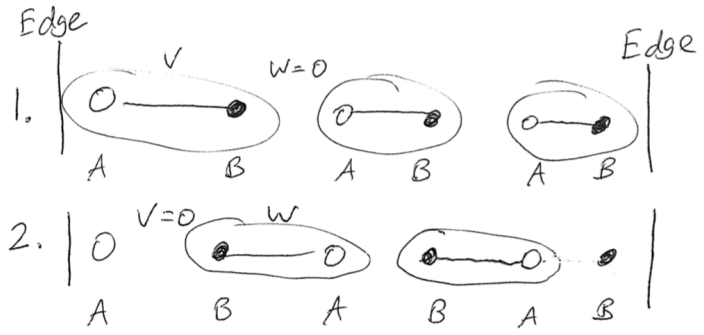


Figure 9: Illustration of edge states in the 'topological' phase of the SSH model, by considering the limiting cases. In case 1, which corresponds to a winding number $W = 0$, there are no edge states. In case 2 where $W = 1$, we see that there are zero energy states localised at each edge.

3.2 Edge states

Before getting into more detail on the symmetry protection, let us look at the bulk-boundary correspondence in the SSH model. As with the Haldane model, we can first make the general observation that if the parameters of the model vary in space, going from a region where $W = 1$ to a region where $W = 0$, then necessarily at the interface between these two regions, we must have a gapless state, as this is the only way the topological index W can change. Notice rather than being a chiral mode, we are now in one less dimension, so we in fact have just a single state localised on the edge. Also notice that again, we can make the argument that the winding number of the vacuum $W = 0$, so a finite lump of SSH material in the 'topological' phase where $W = 1$ must have edge states.

The SSH model is quite nice as it is very easy to visualise these in a simple picture, by going to the limits $w = 0$ or $v = 0$, which is illustrated in Fig. 9. Let us make a few observations:

1. In the limit $v = 0$, which corresponds to the topological phase, $W = 1$, we can clearly see the edge modes, which are localised on a single site which has nothing to hop to.

2. We can also clearly see from the figure that if we applied periodic boundary conditions to the model, the edge modes would disappear, and there would be no difference between cases 1 and 2. While we already made this point for the Haldane model, it is much clearer to see it here.

3. If one goes away from the limit $v = 0$, the edge state will no longer be localised on a single lattice site. However, the topological argument tells us that the edge state must still exist, up to the transition point $v = w$ where the model becomes gapless.

3.3 The protecting symmetry

We now get to the main point of this section, which is to show how a symmetry may force $h_z = 0$. In quantum mechanics, we usually say that a Hamiltonian has a symmetry

represented by a unitary operator U if the commutator

$$[H, U] == HU - UH = 0. \tag{51}$$

This can be simply rearranged to give an alternative form

$$UHU^\dagger = H. \tag{52}$$

Ultimately, this shows that U and H can be diagonalised together, which means that we can partition the eigenstates of H into different sectors with different eigenvalues of U. In fact, this is exactly what Bloch's theorem and writing the Hamiltonian in momentum space is doing if U represents translational symmetry. So translational symmetry has reduced the Hamiltonian from being a $2N \times 2N$ matrix (where N is the number of unit cells in the system) to N 2×2 matrices, labelled by the momentum eigenvalue k. It is clear by analogy that if there were a further symmetry of this type, one could reduce the 2×2 matrix to two scalers, labelled by different eigenvalues of that symmetry, and which have independent dynamics. If this were the case, there would be no chance of any topological quantum number – but more importantly for the present discussion, it cannot possibly force $d_z = 0$, i.e. force certain matrix elements of a Hamiltonian to be zero.

So while unitary symmetries are important for labelling the states of a Hamiltonian, they are useless when it comes to symmetry protected topological states. Symmetry protection requires non-unitary symmetries. The best example of this is time-reversal, which is anti-unitary (it involves complex conjugation), which we will come onto in the next section. Let us introduce another one now though, which is known as a *chiral* symmetry. We say that a Hamiltonian has a chiral symmetry if

$$\Gamma H \Gamma^\dagger = -H, \tag{53}$$

where Γ is unitary **and** Hermition, i.e. satisfies $\Gamma^2 = 1$.[27] The only difference between this and the unitary symmetry is the extra $-$ sign, but this makes all the difference, as it means Γ actually anti-commutes with the Hamiltonian rather than commuting with it, so all the conventional wisdom about good quantum numbers etc... is gone and we are left with something new.

I will leave the interested reader to figure out various consequences of having a chiral symmetry,[28] but let us see it in action for the SSH model, which has such a symmetry, with

$$\Gamma = \sigma_z. \tag{54}$$

This satisfies the requirements of being unitary, and $\sigma_z^2 = 1$; and furthermore, it is easy to show that

$$\sigma_z \sigma_x \sigma_z = -\sigma_x \tag{55}$$

$$\sigma_z \sigma_y \sigma_z = -\sigma_y \tag{56}$$

$$\sigma_z \sigma_z \sigma_z = +\sigma_z, \tag{57}$$

[27] The reason for the requirement that $\Gamma^2 = 1$ is simply that if it didn't, one could decompose the chiral symmetry into one in which $\Gamma^2 = 1$ and a normal unitary symmetry. So assuming one has already dealt with all conventional unitary symmetries in the model, the chiral symmetry must have $\Gamma^2 = 1$. This isn't terribly important for the current discussion, but is quite important if you work in the field, so I will leave the interested students to read more about it elsewhere.

[28] A fun exercise for example is to show that the chiral symmetry implies the spectrum must be symmetric. Read the literature to find out more.

so that

$$\sigma_z H \sigma_z = -H \tag{58}$$

only if H doesn't contain any terms proportional to σ_z, which is exactly what we needed to define the winding number as a topological invariant. Hence the SSH model does indeed have a chiral symmetry; and any perturbation that respects this symmetry will not change the topological index of the state or the presence of zero-energy edge states.

The definition of the chiral symmetry is mathematically very nice, but physically completely opaque. There is another way of looking at it however, as it turns out that chiral symmetry is the same as something called sublattice symmetry. To see this, define two operators

$$P_A = \frac{1}{2}(1+\Gamma), \quad P_B = \frac{1}{2}(1-\Gamma). \tag{59}$$

Elementary algebra shows that these satisfy $P_A + P_B = 1$ and $P_A P_B = 0$, which is the defining property of projection operators – i.e. P_A projects the Hamiltonian onto one subspace A and P_B projects the Hamiltonian onto the rest of the states, B. You can verify that for the case of the SSH model, this is exactly projection onto sublattices A and B defined above, although in other models it may be more abstract than that. The chiral symmetry then is equivalent to the statement

$$P_A H P_A = P_B H P_B = 0, \tag{60}$$

i.e. the Hamiltonian only has terms that link sublattices A and B, and no matrix elements between states on the same sublattice.

Let us summarise what we have learned so far:

1. Unitary symmetries are irrelevant when discussing topological protection
2. There is another sort of symmetry possible, known as chiral symmetry or sublattice symmetry defined above, which are equivalent. The SSH model is an example of a topological insulator protected by the chiral/sublattice symmetry.

We have seen explicitly how this works in the SSH model, so let us now proceed to look at another sort of non-unitary symmetry – time reversal.

4 Time reversal symmetry and the quantum spin Hall effect

The field of topological insulators really emerged when the first time-reversal symmetric topological state was found in 2005 by Kane and Mele. Their model added a spin-orbit term to the graphene Hamiltonian which we will discuss shortly. Spin-orbit tends to be very weak in light elements such as carbon, so this state was completely impractical to see in experiments – however it showed what ingredients were necessary and paved the way for more realistic materials predictions, leading to the experimental verification in Mercury Telluride quantum wells just a few years later. Let us first examine the model, then we will discuss the role of time-reversal symmetry in the topological properties.

4.1 The Kane-Mele model

The Kane-Mele model adds a spin-orbit term to the graphene Hamiltonian (a tight binding model on a hexagonal lattice), and has the Hamiltonian

$$H = H_{\text{graphene}} + i\lambda_{SO} \sum_{\text{n.n.n}} \nu_{ij} c_i^\dagger S^z c_j. \tag{61}$$

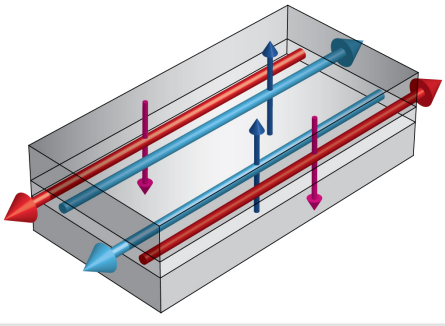


Figure 10: Illustration of helical edge states. The spin-up electrons have one chirality; the spin-down electrons have the opposite chirality. Figure credit: RIKEN website.

In this expression, the sum is over next-nearest neighbors ij, the parameter $\nu_{ij} = \pm 1$ depending on which way round the loop the bond is going, and S^z is the z projection of the spin of the electron. You have already seen this Hamiltonian in the mesoscopic lectures, derived using symmetry considerations.

You have also already seen this Hamiltonian in these lecture notes! Let us consider a spin-up electron, so $S^z = +1/2$. Then comparing with Fig. 3, we see that the Hamiltonian (61) is exactly the Haldane model with $M = 0$, $\phi = \pi/2$,[29], and $t_2 = \lambda_{SO}/2$. Similarly, if we look at a spin-down electron with $S_z = -1/2$, then we have exactly the same thing, but with $\phi = -\pi/2$.

We can therefore read the topological properties straight off Fig. 4 – the spin up electrons have Chern number $+1$, and the spin-down electrons have Chern number -1. Note that the total Chern number for the system is the sum of these two, which is zero – seemingly implying that the state is not topological in nature. Indeed, the Chern number changes sign under time reversal, so if the system is time-reversal-invariant, then the total Chern number must be zero. Before understanding the role topology plays in this system however, let us look at the edge states.

4.2 Helical edge states

We have seen that the Kane-Mele model is two copies of the Haldane model, with $C = 1$ for spin-up electrons, and $C = -1$ for spin-down electrons. We also know from our discussion of the Haldane model that there are gapless topologically protected edge states with chirality C.[30] This is illustrated in Fig. 10, of which a few comments can be made:

1. Such edge states are known as *helical*, in analogy with helicity in high energy physics which means that the spin projection is fixed relative to the momentum direction. To be complete, whether we get positive or negative helicity depends on the sign of λ_{SO} in the Kane-Mele model. Like chiral edge states, these don't naturally occur in non-interacting one-dimensional models.

2. This also illustrates nicely the time-reversal symmetry present in the model – if one changes the direction of time, then one negates the momentum, and also flips

[29] so $e^{i\phi} = i$, which is the i in front of the spin-orbit term in the Hamiltonian

[30] Strictly speaking, there are $|C|$ edge modes with chirality given by the sign of C, but if we limit ourselves to $C = \pm 1$, this is the same thing.

the spin.

Now we come to the role of topological protection of helical edge states. Let us first consider the Kane-Mele model as it is written, where S_z is conserved. This means that one cannot flip and up-spin into a down-spin, or vice versa. Then one may look only at one-spin projection at a time – and as soon as one does this, we have a non-zero Chern number in the bulk, and a chiral edge state, which we have already discussed. Thus, any perturbation to the model that also conserves S_z can not gap the edge sttaes.

This actually cements a point we laboured over earlier in the SSH model. A conservation law comes from a symmetry,[31] and in the case of S_z conservation, the symmetry is (spin) rotation symmetry about the z axis, which is a *normal unitary symmetry.* Hence what has the unitary symmetry done? The Kane-Mele Hamiltonian is a 4×4 matrix in k space, because we have the 2×2 sublattice structure of graphene that we already saw in the Haldane model, times an addition 2×2 structure for spin. The unitary symmetry that leads to S_z conservation however splits this into two separate 2×2 Hamiltonians that we can *analyse independently.* Hence the unitary symmetries are very important in the analysis of topology because they get rid of degeneracies that can be labelled by good quantum numbers. However, once you get to the minimum block size of Hamiltonian, they don't do any more. The Kane-Mele model is an excellent example to see this in action.

Conversely,. suppose we did not have S_z conservation, and furthermore there were no other symmetries in the system. Then the relevant topological index is the total Chern number (which cannot be subdivided into independent contributions any more) which is 0, so the system is not topological. One can also see this from the edge states – as soon as we don't care about the spin-projection, they are just ordinary one-dimensional modes – which can be localised by disorder. Hence in both ways of looking at it, the system is not topological.

However, in the SSH model, we have already seen if one has an additional non-unitary symmetry in the model, then one can get a symmetry protected topological state. The relevant symmetry in the SSH model was chiral/sub-lattice symmetry. In the Kane-Mele model, this role is played by time reversal symmetry, which we now discuss.

4.3 The role of time reversal symmetry

Let us start by writing the single particle time dependent Schrödinger equation of a particle in a scaler and vector potential:

$$i\partial_t \psi(r,t) = \left[(-i\nabla - eA(r))^2 + V(r)\right] \psi(r,t) \tag{62}$$

Taking the complex conjugate of this equation gives us

$$-i\partial_t \psi(r,t)^* = \left[(-i\nabla + eA(r))^2 + V(r)\right] \psi(r,t)^*, \tag{63}$$

which is basically the same, except that the vector potential has changed sign, and the derivative with respect to time has changed sign. This is exactly what you would expect under time-reversal – the vector potential is associated with magnetic fields which change sign under time reversal, and changing the direction of time will change the sign of ∂_t. You've probably seen a more formal version of this in undergraduate quantum mechanics, but this is a quick demonstration that shows the time reversal operator in

[31] Noether's theorem classically, which was extended to quantum mechanics by Ward and collaborators, but that is a different story entirely.

quantum mechanics involves complex conjugation, which is known as an anti-unitary operator.

In addition to complex conjugation, if the wavefunction has an internal degree of freedom then time reversal may also involve a unitary transformation on this internal state. For example, we know for a spin-1/2 system that

$$T|\uparrow\rangle = |\downarrow\rangle, \quad T|\downarrow\rangle = -|\uparrow\rangle, \tag{64}$$

hence the full time reversal operator T for a spin-1/2 system may be written as

$$T = -\sigma_y K, \tag{65}$$

where σ_y acts on the actual spin degrees of freedom (and so should not be confused with our use of the Pauli matrices in an arbitrary two-band system for most of these notes), and K is the complex conjugate operator. A Hamiltonian is then invariant under time reversal if

$$THT^{-1} = H \tag{66}$$

You may find it curious in this example that if one applies the time reversal symmetry operator twice, $T^2 = -1$, one does not get back to the original state, but obtains an extra negative sign. This may at first seem paradoxical, because surely if you reverse time, then reverse it back again, you haven't done anything. However, in quantum mechanics, if you multiply the wavefunction by an arbitrary phase, you also haven't done anything. So there is absolutely nothing contradictory about T^2 giving a phase. It is more involved[32] to show that the phase must be ± 1, and that $T^2 = +1$ corresponds to bosons (or integer spin particles), while $T^2 = -1$ corresponds to fermions (or half-integer spin particles). The main thing to be aware of is that there are two different possibilities for time reversal symmetry, and when it comes to topological classification, this matters.

4.3.1 Kramers pairs

Let us consider a Hamiltonian that is time reversal symmetric, with $T^2 = -1$[33] Kramers theorem states that all eigenstates of the Hamiltonian are doubly degenerate, in time reversal pairs known as Kramers doublets or Kramers paris.

This is easy to prove. Suppose there is an eigenstate of the Hamiltonian, $|\chi\rangle$ that is non-degenerate. Then applying the time reversal operator to it can give at most a multiplying factor c

$$T|\chi\rangle = c|\chi\rangle. \tag{67}$$

Applying again gives

$$T^2|\chi\rangle = |c|^2|\chi\rangle, \tag{68}$$

where we get $|c|^2 = c^*c$ rather than just c^2 because the time-reversal operator involves complex conjugation. However, $T^2 = -1$, which means that $|c|^2 = -1$, which is not possible. Hence the state must be degenerate. Furthermore, one can prove[34] that

$$\langle T\chi|\chi\rangle = 0, \tag{69}$$

i.e. the time-reversed state is *orthogonal* to the original state.

[32] And comes from relativistic quantum mechanics – I find Weinberg's book on the quantum theory of fields part I a good reference if you want to go deeper into this.

[33] It is trivial, but nevertheless a useful exercise to see why the theorem doesn't hold if $T^2 = +1$.

[34] Do it!

4.3.2 $\mathbf{Z}_2$ topology

Let us now return to our helical edge states from Fig. 10 and see what all of this discussion of time-reversal symmetry and Kramers pairs has to do with the edge states. We have already seen that if S_z is conserved, then these states are almost trivially topologically protected because they can be separated into independent chiral modes. We have also seen that if there is no symmetry, then these states have no topological protection and can be gapped out by e.g. disorder. Let us now however consider an intermediate case, where S_z isn't necessarily conserved,[35] however time-reversal symmetry is present.

The Kane-Mele model is a spin-1/2 system, so $T^2 = -1$ and we have Kramer's theorem, which in particular on the edge tells us that $|\uparrow, k\rangle$ and $|\downarrow, -k\rangle = |\uparrow, k\rangle$ are orthogonal. This means that no perturbation to the model (e.g. an impurity or disorder) that respects time reversal symmetry can scatter between these two states. In particular, this means that one may have no *elastic* scattering from impurities on the edge.[36] Hence so long as time reversal symmetry is present, the helical edge mode is topologically protected – and hence by the bulk-boundary correspondence in reverse, there should be some non-zero bulk topological index.

This can not be given by the Chern number of each spin projection however. To begin with, while we are currently viewing the Kane-Mele model as it is, with perturbations that break S_z conservation (but respect time reversal symmetry), this isn't necessarily the case. One could certainly write a model where there is no way of splitting it into an S_z conserving part plus a small perturbation – at which point S_z is not a good quantum number, so cannot be used to label states, and by extension, one cannot restrict to a single value of S_z. However, there is an even more fundamental reason why the Chern number of each spin projection is not a topological invariant in this case, which can be easily seen by considering two copies of the Kane-Mele model. In this case, one has two copies of the helical states – and while Kramers theorem prohibits elastic scattering from one state into its Kramers partner; there is nothing to stop it from elastically scattering from one state to the opposite direction in the other copy of the helical state. Hence this case is like the one without symmetry – and the edge states can be gapped.

It is not difficult to extend this argument to show that any even number of helical edges can be gapped; however for any odd number, there is always one left over that can't localise under time-reversal symmetric disorder. This means that there is a topological index that takes only two values, which is known as a $\mathbf{Z}_2$ invariant, in contrast to the Chern number which takes any integer value, so is known as a $\mathbf{Z}$ invariant.

So how can one calculate this $\mathbf{Z}_2$ invariant, which we will call ν, for a time-reversal symmetric[37] insulating state?

1. If S_z is conserved, then one can obtain it from the Chern number, $\nu = C_{\uparrow,\downarrow}$ modulo 2. It doesn't matter which spin projection is taken because $C_\uparrow = C_\downarrow$ for a time reversal symmetric system. Notice that by saying S_z is conserved, one is imagining a Hamiltonian like the Kane-Mele model which conserves S_z, but with some perturbations that don't conserve S_z. As we already discussed, if one only considers perturbations which conserve S_z, then ν is an irrelevant quantity, as one should use the unitary symmetry to split up the Hamiltonian further into independent sectors which can be indexed by the Chern number.

[35] Which is more likely to be the case – the Kane-Mele model is idealised in the way it writes spin-orbit, as in a general (slightly disordered) system, there will be no axis where spin is conserved.

[36] Inelastic scattering is still possible, although usuall weak, but this goes beyond the non-interacting paradigm.

[37] with $T^2 = -1$ – not always explicitly said when they are discussed, but important.

2. If inversion symmetry is present, then one can show that ν is equal to the product of the parity of the wavefunctions at the high symmetry points.

3. In general, there are about 10 different way (that I'm aware of) to calculate ν. None of them are particularly nice, and therefore I leave reading about them in the literature to those who want or need to know more technical details.

5 The periodic table of topological insulators and superconductors

We've now seen examples of all of the ingredients needed to understand non-interacting topological bandstructures in general. We've seen examples of $\mathbf{Z}$ and $\mathbf{Z}_2$ topological indices; and we've also seen the role of symmetries and exactly what is meant by symmetry protected topological state. It turns out that one can classify all possible topological band structures based on these symmetries – which was first done by Schnyder, Ryu, Furusaki and Ludwig, and independently by Kitaeev in 2008/2009.

5.1 Non-unitary symmetries

We have already seen that unitary symmetry

$$UHU^{-1} = H \tag{70}$$

splits the Hilbert space into independent parts, but doesn't do anything towards symmetry protected topological states. We have also already seen two examples of non-unitary symmetries:

1. Chiral, or sub-lattice symmetries, characterised by a unitary operator Γ with $\Gamma^2 = 1$, with
$$\Gamma H\Gamma^{-1} = -H \tag{71}$$

2. Time reversal symmetry, characterised by $T = \tau K$ where τ is unitary and K is complex conjugation, with
$$THT{-}1 = H \tag{72}$$

3. One can also combine these two cases we have already seen, to get a third possibility for a non-unitary symmetry P. Like time reversal, this involves a unitary operator $\tilde{\tau}$ and complex conjugation $P = \tilde{\tau}K$, and like chiral symmetry, has an extra minus in the symmetry requirement:
$$PHP{-}1 = H \tag{73}$$
In physical examples, one can relate this to a particle hole symmetry – this intuitively makes sense because changing a particle to a hole should involve complex conjugation, but it also must be different from time reversal. Read the literature to find more details about this though. Like time reversal, P^2 can be plus or minus one.

It turns out that these are the only non-unitary symmetries that one can use to characterise Hamiltonians, which was known since the work of Wigner and Dyson on random matrix theory. The possibilities are all listed in Table. 1, and it is worth making a few comments:

		T	P	C	1D	2D	3D
Standard	A	0	0	0	-	$\mathbf{Z}$	-
	AI	+1	0	0	-	-	-
	AII	−1	0	0	-	$\mathbf{Z}_2$	$\mathbf{Z}_2$
Chiral	AIII	0	0	1	$\mathbf{Z}$	-	$\mathbf{Z}$
	BDI	+1	+1	1	$\mathbf{Z}$	-	-
	CII	−1	−1	1	$\mathbf{Z}$	-	$\mathbf{Z}_2$
Bogolyubov de Genne	D	0	+1	0	$\mathbf{Z}_2$	$\mathbf{Z}$	-
	C	0	−1	0	-	$\mathbf{Z}$	-
	DIII	−1	+1	1	$\mathbf{Z}_2$	$\mathbf{Z}_2$	$\mathbf{Z}$
	CI	+1	−1	1	-	-	$\mathbf{Z}$

Table 1: Classification of all possible non-interacting topological states in one, two, and three dimensions. The columns T, P, and C tell us whether a time-reversall, particle-hole, and/or chiral symmetry is present, and in the first two cases what the square of the operator gives. The labels correspond to the Wigner-Dyson classifcation of random matrices. A dash means no topological bands are possible in this symmetry class and number of dimensions.

1. While we introduced P as a combination of T and Γ, it is much more common to say that $\Gamma = TP$.[38] This means that if one has time-reversal and particle-hole symmetries, one must also have chiral symmetry. If one has one of particle-hole or time-reversal, then one can't have chiral symmetry. The only case where chiral symmetry is important is if one doesn't have either time-reversal or particle-hole symmetry – because in this case, one may still have the product of them. This is why there are ten classes – $T^2 = 0, \pm 1$ is three cases, times $P^2 = 0, \pm 1$ makes 9, but the case where $T = P = 0$ splits into two cases as we may have $C = 0, 1$.

2. The symmetries may be more abstract than we have defined them – one may have a system where time-reversal doesn't make physical sense, but one still mathematically has time-reversal like symmetry, as defined in Eq. (72) above, and the same for the others.

3. Related to this last point, while each of the above symmetries may be in some abstract Hilbert space, it is not possible to have more than one e.g. time-reversal like symmetry. This is because if you do, one can break it down into a single time-reversal like symmetry and another unitary symmetry, which should be analysed accordingly to reduce the Hilbert space.

4. While particle-hole symmetry is what one gets in superconductors, in this abstract sense, this is only true if $P^2 \neq T^2$. This is because if $P^2 = T^2$, then one may induce the particle-hole symmetry from the presence of time-reversal and chiral symmetries, so the presence of P symmetry doesn't necessarily mean superconductivity.

5.2 Classification

Now that we have classified all possible Hamiltonians, we can work out whether there is a possible topological index to them or not – which is shown in 1, 2 and 3 dimensions in

[38] Can you show why this is necessarily the case?

Table. 1. Before discussing the table, let us get some insight into how this is worked out, with a couple of examples we have already seen in the SSH model. We limit ourselves to one-dimension so that the Brillouin zone is a circle, and we will also limit ourselves to two-band systems (i.e. Hamiltonian's that are 2×2 in k space).

1. If we have chiral symmetry present (i.e. classes AIII, BDI and CII in the table), then the space of Hamiltonians is a circle. We therefore map a circle onto a circle, which can be characterised by the winding number, which is a $\mathbf{Z}$ topological invariant.

2. On the other hand, if we don't have chiral symmetry (class A in the table), then the mapping is from a circle to a sphere. We have already discussed that all circles on a sphere can be continuously deformed to all other circles – hence there is no topological invariant.

While these are simple examples, they illustrate how the table is made – one knows the shape of the Brillouin zone in any number of dimensions, one works out the shape of the space of Hamiltonians[39], and then use known results from differential topology to see if there is a topological invariant or not, and if so, whether it is $\mathbf{Z}$ or $\mathbf{Z}_2$.

There are a number of things you could observe about the table – for example, in any number of dimensions, there are precisely 3 of the 10 universality classes supporting a $\mathbf{Z}$ topology, and a further 2 supporting $\mathbf{Z}_2$ topology. Another thing people like to mention is something known as Bott periodicity[40] – it turns out that the classification in $d + 8$ dimensions is identical to that in d-dimensions. For now though, let us limit ourselves to just a few entries, and see what universality classes the models we've looked at so far fall into:

1. The quantum Hall effect (and quantum anomalous Hall effect, e.g. the Haldane model) is class A – with no symmetries at all. The important thing about this class is that there is no topological invariant possible in any odd number of dimensions (including the physical possibilities of one and three). So this table shows that an analogue of the quantum Hall effect in three dimensions is not possible, which is partly the reason why symmetry protected topological states (all of the other universality classes) became so interesting recently.

2. The quantum spin-Hall effect (e.g. the Kane-Mele model) is class AII, with time reversal symmetry (with $T^2 = -1$) and nothing else. One of the most important aspects of this state which we haven't had time to go into at all in these lectures is that it has an analogue in three dimensions. These were seen experimentally is Bismuth Telluride, and I refer you to the literature to explore this further.

3. The SSH model has chiral symmetry, and so may seem like it is universality class AIII – but it is actually BDI. While we didn't discuss it, it also has a time reversal (like) symmetry (and if it has both chiral and time reversal, then it must have particle-hole). I leave it as an exercise for the reader to explicitly work out the T-like and P-like operators under which the SSH Hamiltonian is symmetric.

The first six universality classes in the table can all be written as models of non-interacting fermions. To get the last four requires anomalous terms which don't conserve particle number, which come e.g. from the mean-field description of superconductors,

[39] which is often colloquially given some group structure, but it actually is a symmetric space, a distinction that is probably irrelevant to most of you!

[40] A beautiful mathematical result – but with no clear relevance to condensed matter physics!

hence the term 'Bogolyubov de Genne'. While the topic of topological superconductors is fascinating by itself, there is no time to discuss it in any detail here, however now you know the basics, it will be easy to follow the literature or one of the many excellent review articles.

6 Beyond topological band theory

These lectures have concentrated on topological band theory – so what happens if you go beyond this? A similar question could be: 'what lies beyond the shire'? Ultimately, topology is a huge branch of maths, and people use 'topological' in physics to mean anything that uses this branch of maths. In some sense, a good analogy would be talking about a 'calculus state' – i.e. a state which can be described using calculus, which pretty much only rules out certain discrete models in statistical physics and contains everything else. Let me list just a few of the other places where topology commonly enters condensed matter physics:

1. Topological excitations, such as the vortices in the Kosterlitz-Thouless model.

2. Topological order – you have already met mean field theory, and the characterisation of ordered phases through symmetry breaking and Landau theory. However in some cases, one gets an ordered phase, but with no (local) order parameter, and hence no Landau theory. Like topological band theory, this can either be protected by a symmetry, or not, but it has no non-interacting description. The best example of this is the fractional quantum Hall effect – an entire subject area in its own right!

3. Topological terms – sometimes, rather than talking about topological states, one talks about topological terms in the Hamiltonian. These are simply terms in the Hamiltonian which are topological in nature (i.e. continuously depend on a continuous variable, but have a discrete set out values). There are a limited number of topological terms one can write, the most common being Wess-Zumino, Chern-Simons, or theta-terms. Note that just because the Hamiltonian description has a topological term in it does not say anything about whether the eigenstates of this Hamiltonian can be characterised by a topological index or not (and vice versa, as these notes show).

4. Topological field theory – an extreme example of topological terms in the Hamiltonian, where the Hamiltonian *only* contains topological terms.

5. Topological protection (often associated with qubits or quantum computing) – we already discussed this in the context of topological band theory, but it is much more general. This is back to states, where if they are characterised by some topological index (whether from band theory or not), then small perturbations can't change this. This is clearly very useful in tasks like quantum computing when one wants to make sure small perturbations as one always has in reality can't affect the calculation.

The list could go on, but I think there is enough to make the point. Even if one goes back to topological band theory, there is a lot of current research in understanding what happens to topological invariants when interactions are present.

To summarise, topology is everywhere, and hopefully these notes have given some insight, not just into topological insulators, but also how topological quantum numbers

work in a more general setting, allowing you to dive into the more in-depth reviews of the subject.

7 Further reading

I've made a point not to put too many citations into these lecture notes in order to make them self-contained, and also because you can find references to original articles in any of the lecture notes or reviews cited below. Rather than jumping from these lecture notes direct to the original articles, it is better to read a little more about the subject first, and so what I have listed here is what I think are some of the best pedagogical places to learn about the topic.

7.1 Lecture notes available on the web

1. Asbóth, Oroszlány, Pályi – `http://arxiv.org/abs/1509.02295`. This is the nicest set of lecture notes I have come across, that is entirely self contained. It is also now published as a book in the "Lecture notes in physics" series by springer; the title is "A short course of topological insulators".

2. Kai Sun – `http://www-personal.umich.edu/~sunkai/teaching/Fall_2013/phys620.html`. These are very nice lecture notes on a graduate solid state course at the university of Michigan; much of the course is taken up with topological insulators.

7.2 Books

3. Fradkin, *Field Theories of Condensed Matter in physics*. The second edition has a chapter on topological insulators, which is interesting, however work through some of the lecture notes above first before trying to read this. The book also contains rather a lot on topological phases in general including fractional quantum Hall and spin liquids, that weren't covered here.

4. Wen, *Quantum Field Theory*. This is one of the best books that covers contemporary topics in interacting topological phases, by one of the most active people in field. Very good for further reading, but by no means elementary.

5. Nakahara, *Geometry, topology and physics*. If you want to understand the mathematical background of topology, this is one of the few books that goes through it at a sensible level. Although it has physicists in mind, it is (quite naturally) very mathematical in nature, and if you haven't done a maths course in analysis which includes the construction of the real numbers, it would be well worth doing so before trying to read this book.

7.3 Review articles

6. Hasan and Kane, *Colloquium: Topolological insulators*, Rev. Mod. Phys. **82**, 3045 (2010).

7. Qi and Zhang, *Topological insulators and superconductors*, Rev. Mod. Phys. **83**, 1057 (2011).

8. Liu, Zhang, Qi, *the Quantum Anomalous Hall effect*, `http://arxiv.org/abs/1508.07106`.

9. Rachel, *Interacting Topological Insulators: a review*, `http://arxiv.org/abs/1804.10656`.

8 Workshop questions

Chern Number

1. In its simplest form, the Hamiltonian of a $p + \mathrm{i}p$ superconductor in two spatial dimensions may be written as

$$\mathcal{H} = \begin{pmatrix} -2t\cos k_x - 2t\cos k_y - \mu & \Delta(\sin k_x - \mathrm{i}\sin k_y) \\ \Delta(\sin k_x + \mathrm{i}\sin k_y) & 2t\cos k_x + 2t\cos k_y + \mu \end{pmatrix}$$

   where the gap $\Delta > 0$, the hopping $t > 0$. The parameter μ is a control parameter, and is not physically the chemical potential. You should have already come across Hamiltonians like this in the quantum fluids course; you will meet them again in the unconventional superconductivity course. For now, you do not need to worry where the Hamiltonian comes from. The Brillouin zone is defined as $|k_{x,y}| \leq \pi$.

   By calculating the Chern index, show under what conditions the ground state of this Hamiltonian is topological.

2. Consider a general two band lattice Hamiltonian, which can be expressed as a 2×2 matrix:

$$\mathcal{H} = \underline{h}(\underline{k}) \cdot \underline{\sigma}$$

   where $\underline{\sigma}$ is the vector of Pauli matrices.

   If $\underline{u}(\underline{k})$ are the appropriate eigenvectors, then the Berry connection $\underline{A}(\underline{k}) = i\underline{u}^* \cdot \nabla \underline{u}$ and the Berry curvature is

$$F(\underline{k}) = \frac{\partial A_y}{\partial k_x} - \frac{\partial A_x}{\partial k_y}.$$

   Show that for this Hamiltonian, the Berry curvature as defined above may alternatively be written as

$$F(\underline{k}) = -\frac{1}{2}\underline{d} \cdot \left(\frac{\partial \underline{d}}{\partial k_x} \times \frac{\partial \underline{d}}{\partial k_y} \right).$$

   where the unit vector field $\underline{d} = \underline{h(\underline{k})}/|\underline{h(\underline{k})}|$.

   *H*int: There are many ways to do this using fancy maths; however I favour for this case the brute force approach: calculate the eigenvectors of the Hamiltonian and from them the Berry curvature; and show it is equivalent to the expression above.

   This is a very important relation as it allows one to directly see the meaning of the Chern number

$$C = \frac{1}{4\pi} \oint_{\mathrm{BZ}} \underline{d} \cdot \left(\frac{\partial \underline{d}}{\partial k_x} \times \frac{\partial \underline{d}}{\partial k_y} \right) d^2\underline{k}.$$

   The Hamiltonian $\underline{d}(\underline{k})$ is a mapping from the surface of a taurus (the Brillouin zone) to the surface of a sphere (the Hamiltonian, which in this case is defined by the vector $\underline{(d)}$. The Jacobian of this mapping is just the Berry curvature as defined above (Conceptual questions: can you see this?), meaning that the integral is the surface area of the sphere (4π) times the number of times the field $\underline{d}$ wraps around the sphere. This is a higher dimensional analogue of the simple winding number.

3. Another model that exhibits the quantum anomalous Hall effect is given, in the notation of the last question, by

$$\underline{d}(\underline{k}) = \begin{pmatrix} A \sin k_x \\ -A \sin k_y \\ M + 2B(2 - \cos k_x - \cos k_y) \end{pmatrix}.$$

Roughly speaking, this is half of the 'BHZ' model for the quantum spin-Hall effect from lecture 2. The model was also discussed in Qi, Wu and Yang, Phys. Rev. B **74** 085308.

Analyse this model, and state under which conditions it is topological.

4. Show that the Hall conductance σ_{xy} in the quantum anomalous Hall effect is related to the Chern number as

$$\sigma_{xy} = \frac{e^2}{h} C.$$

This question is more advanced, and for the 'experts' only. The usual way to do it is via the Kubo formula for conductance; the paper cited in the previous question goes through this, as well as Haldane's 1988 paper. However there are probably other more conceptual (but less rigorous) ways of getting this result through an analogue of the Laughlin construction for regular quantum Hall.

Edge modes

5. Consider the one-dimensional continuum Hamiltonian

$$\mathcal{H} = -\mathrm{i} v_F \sigma^x \partial_x + m(x) \sigma^z$$

Consider a domain wall in the mass $m(x)$, i.e. $m(x)$ is some (monotonic) function that goes to $-m_0$ at $x \to -\infty$ to $+m_0$ at $x =\to +\infty$. Show that there is a solution of the Hamiltonian with zero energy, that is localised around the domain wall in the mass. (This is the Jackiw-Rebbi model, which is in turn a continuum limit of the Su-Schrieffer-Heeger model).

6. Consider the 'half-BHZ' model from question 3 on a strip geometry. By numerically diagonalising the Hamiltonian matrix, investigate how the edge states appear as the parameters in the model are tuned through one of the topological transitions.