

The Quantum Hall Effect

TIFR Infosys Lectures

David Tong

*Department of Applied Mathematics and Theoretical Physics,
Centre for Mathematical Sciences,
Wilberforce Road,
Cambridge, CB3 0BA, UK*

*<http://www.damtp.cam.ac.uk/user/tong/qhe.html>
d.tong@damtp.cam.ac.uk*

Recommended Books and Resources

There are surprisingly few dedicated books on the quantum Hall effect. Two prominent ones are

- Prange and Girvin, “*The Quantum Hall Effect*”

This is a collection of articles by most of the main players circa 1990. The basics are described well but there’s nothing about Chern-Simons theories or the importance of the edge modes.

- J. K. Jain, “*Composite Fermions*”

As the title suggests, this book focuses on the composite fermion approach as a lens through which to view all aspects of the quantum Hall effect. It has many good explanations but doesn’t cover the more field theoretic aspects of the subject.

There are also a number of good multi-purpose condensed matter textbooks which contain extensive descriptions of the quantum Hall effect. Two, in particular, stand out:

- Eduardo Fradkin, *Field Theories of Condensed Matter Physics*
- Xiao-Gang Wen, *Quantum Field Theory of Many-Body Systems: From the Origin of Sound to an Origin of Light and Electrons*

Several excellent lecture notes covering the various topics discussed in these lectures are available on the web. Links can be found on the course webpage: <http://www.damtp.cam.ac.uk/user/tong/qhe.html>.

Contents

1. The Basics	5
1.1 Introduction	5
1.2 The Classical Hall Effect	6
1.2.1 Classical Motion in a Magnetic Field	6
1.2.2 The Drude Model	7
1.3 Quantum Hall Effects	10
1.3.1 Integer Quantum Hall Effect	11
1.3.2 Fractional Quantum Hall Effect	13
1.4 Landau Levels	14
1.4.1 Landau Gauge	18
1.4.2 Turning on an Electric Field	21
1.4.3 Symmetric Gauge	22
1.5 Berry Phase	27
1.5.1 Abelian Berry Phase and Berry Connection	28
1.5.2 An Example: A Spin in a Magnetic Field	32
1.5.3 Particles Moving Around a Flux Tube	35
1.5.4 Non-Abelian Berry Connection	38
 2. The Integer Quantum Hall Effect	 42
2.1 Conductivity in Filled Landau Levels	42
2.1.1 Edge Modes	44
2.2 Robustness of the Hall State	48
2.2.1 The Role of Disorder	48
2.2.2 The Role of Gauge Invariance	51
2.2.3 An Aside: The Kubo Formula	54
2.2.4 The Role of Topology	57
2.3 Particles on a Lattice	61
2.3.1 TKNN Invariants	61
2.3.2 The Chern Insulator	65
2.3.3 Particles on a Lattice in a Magnetic Field	67

3. The Fractional Quantum Hall Effect	74
3.1 Laughlin States	76
3.1.1 The Laughlin Wavefunction	76
3.1.2 Plasma Analogy	80
3.1.3 Toy Hamiltonians	83
3.2 Quasi-Holes and Quasi-Particles	85
3.2.1 Fractional Charge	88
3.2.2 Introducing Anyons	90
3.2.3 Fractional Statistics	92
3.2.4 How to Detect an Anyon	97
3.2.5 Ground State Degeneracy and Topological Order	100
3.3 Other Filling Fractions	101
3.3.1 The Hierarchy	102
3.3.2 Composite Fermions	104
3.3.3 The Half-Filled Landau Level	109
3.3.4 Wavefunctions for Particles with Spin	113
 4. Non-Abelian Quantum Hall States	 117
4.1 Life in Higher Landau Levels	117
4.2 The Moore-Read State	119
4.2.1 Quasi-Holes	121
4.2.2 Majorana Zero Modes	126
4.2.3 Read-Rezayi States	132
4.3 The Theory of Non-Abelian Anyons	134
4.3.1 Fusion	135
4.3.2 The Fusion Matrix	138
4.3.3 Braiding	142
4.3.4 There is a Subject Called Topological Quantum Computing	145
 5. Chern-Simons Theories	 146
5.1 The Integer Quantum Hall Effect	147
5.1.1 The Chern-Simons Term	149
5.1.2 An Aside: Periodic Time Makes Things Hot	151
5.1.3 Quantisation of the Chern-Simons level	153
5.2 The Fractional Quantum Hall Effect	156
5.2.1 A First Look at Chern-Simons Dynamics	157
5.2.2 The Effective Theory for the Laughlin States	159
5.2.3 Chern-Simons Theory on a Torus	164

5.2.4	Other Filling Fractions and K -Matrices	166
5.3	Particle-Vortex Duality	170
5.3.1	The XY -Model and the Abelian-Higgs Model	170
5.3.2	Duality and the Chern-Simons Ginzburg-Landau Theory	174
5.3.3	Composite Fermions and the Half-Filled Landau Level	180
5.4	Non-Abelian Chern-Simons Theories	184
5.4.1	Introducing Non-Abelian Chern-Simons Theories	184
5.4.2	Canonical Quantisation and Topological Order	186
5.4.3	Wilson Lines	188
5.4.4	Chern-Simons Theory with Wilson Lines	192
5.4.5	Effective Theories of Non-Abelian Quantum Hall States	200
6.	Edge Modes	201
6.1	Laughlin States	201
6.1.1	The View from the Wavefunction	201
6.1.2	The View from Chern-Simons Theory	203
6.1.3	The Chiral Boson	208
6.1.4	Electrons and Quasi-Holes	210
6.1.5	Tunnelling	215
6.2	The Bulk-Boundary Correspondence	217
6.2.1	Recovering the Laughlin Wavefunction	217
6.2.2	Wavefunction for Chern-Simons Theory	222
6.3	Fermions on the Boundary	226
6.3.1	The Free Fermion	226
6.3.2	Recovering the Moore-Read Wavefunction	229
6.4	Looking Forwards: More Conformal Field Theory	231

Acknowledgements

These lectures were given in TIFR, Mumbai. I'm grateful to the students, postdocs, faculty and director for their excellent questions and comments which helped me a lot in understanding what I was saying.

To first approximation, these lecture notes contain no references to original work. I've included some footnotes with pointers to review articles and a handful of key papers. More extensive references can be found in the review articles mentioned earlier, or in the book of reprints, "*Quantum Hall Effect*", edited by Michael Stone.

My thanks to everyone in TIFR for their warm hospitality. Thanks also to Bart Andrews for comments and typo-spotting and to Steve Kivelson for a number of illuminating comments. These lecture notes were written as preparation for research funded by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013), ERC grant agreement STG 279943, "Strongly Coupled Systems".

Magnetic Scales

$$\text{Cyclotron Frequency: } \omega_B = \frac{eB}{m}$$

$$\text{Magnetic Length: } l_B = \sqrt{\frac{\hbar}{eB}}$$

$$\text{Quantum of Flux: } \Phi_0 = \frac{2\pi\hbar}{e}$$

$$\text{Hall Resistivity: } \rho_{xy} = \frac{2\pi\hbar}{e^2} \frac{1}{\nu}$$

1. The Basics

1.1 Introduction

Take a bunch of electrons, restrict them to move in a two-dimensional plane and turn on a strong magnetic field. This simple set-up provides the setting for some of the most wonderful and surprising results in physics. These phenomena are known collectively as the *quantum Hall effect*.

The name comes from the most experimentally visible of these surprises. The Hall conductivity (which we will define below) takes quantised values

$$\sigma_{xy} = \frac{e^2}{2\pi\hbar} \nu$$

Originally it was found that ν is, to extraordinary precision, integer valued. Of course, we're very used to things being quantised at the microscopic, atomic level. But this is something different: it's the quantisation of an emergent, macroscopic property in a dirty system involving many many particles and its explanation requires something new. It turns out that this something new is the role that topology can play in quantum many-body systems. Indeed, ideas of topology and geometry will be a constant theme throughout these lectures.

Subsequently, it was found that ν is not only restricted to take integer values, but can also take very specific rational values. The most prominent fractions experimentally are $\nu = 1/3$ and $\nu = 2/5$ but there are many dozens of different fractions that have been seen. This needs yet another ingredient. This time, it is the interactions between electrons which result in a highly correlated quantum state that is now recognised as a new state of matter. It is here that the most remarkable things happen. The charged particles that roam around these systems carry a fraction of the charge of the electron, as if the electron has split itself into several pieces. Yet this occurs despite the fact that the electron is (and remains!) an indivisible constituent of matter.

In fact, it is not just the charge of the electron that fractionalises: this happens to the “statistics” of the electron as well. Recall that the electron is a fermion, which means that the distribution of many electrons is governed by the Fermi-Dirac distribution function. When the electron splits, so too does its fermionic nature. The individual constituents are no longer fermions, but neither are they bosons. Instead they are new entities known as *anyons* which, in the simplest cases, lie somewhere between bosons and fermions. In more complicated examples even this description breaks down: the resulting objects are called *non-Abelian anyons* and provide physical embodiment of the kind of non-local entanglement famous in quantum mechanics.

Because of this kind of striking behaviour, the quantum Hall effect has been a constant source of new ideas, providing hints of where to look for interesting and novel phenomena, most of them related to the ways in which the mathematics of topology impinges on quantum physics. Important examples include the subject of topological insulators, topological order and topological quantum computing. All of them have their genesis in the quantum Hall effect.

Underlying all of these phenomena is an impressive theoretical edifice, which involves a tour through some of the most beautiful and important developments in theoretical and mathematical physics over the past decades. The first attack on the problem focussed on the microscopic details of the electron wavefunctions. Subsequent approaches looked at the system from a more coarse-grained, field-theoretic perspective where a subtle construction known as Chern-Simons theory plays the key role. Yet another perspective comes from the edge of the sample where certain excitations live that know more about what’s happening inside than you might think. The main purpose of these lectures is to describe these different approaches and the intricate and surprising links between them.

1.2 The Classical Hall Effect

The original, classical Hall effect was discovered in 1879 by Edwin Hall. It is a simple consequence of the motion of charged particles in a magnetic field. We’ll start these lectures by reviewing the underlying physics of the Hall effect. This will provide a useful background for our discussion of the quantum Hall effect.

Here’s the set-up. We turn on a constant magnetic field, \mathbf{B} pointing in the z -direction. Meanwhile, the electrons are restricted to move only in the (x, y) -plane. A constant current I is made to flow in the x -direction. The *Hall effect* is the statement that this induces a voltage V_H (H is for “Hall”) in the y -direction. This is shown in the figure to the right.

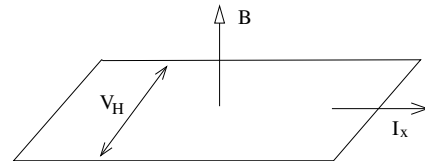


Figure 1: The classical Hall effect

1.2.1 Classical Motion in a Magnetic Field

The Hall effect arises from the fact that a magnetic field causes charged particles to move in circles. Let’s recall the basics. The equation of motion for a particle of mass m and charge $-e$ in a magnetic field is

$$m \frac{d\mathbf{v}}{dt} = -e\mathbf{v} \times \mathbf{B}$$

When the magnetic field points in the z -direction, so that $\mathbf{B} = (0, 0, B)$, and the particle moves only in the transverse plane, so $\mathbf{v} = (\dot{x}, \dot{y}, 0)$, the equations of motion become two, coupled differential equations

$$m\ddot{x} = -eB\dot{y} \quad \text{and} \quad m\ddot{y} = eB\dot{x} \quad (1.1)$$

The general solution is

$$x(t) = X - R \sin(\omega_B t + \phi) \quad \text{and} \quad y(t) = Y + R \cos(\omega_B t + \phi) \quad (1.2)$$

We see that the particle moves in a circle which, for $B > 0$, is in an anti-clockwise direction. The centre of the circle, (X, Y) , the radius of the circle R and the phase ϕ are all arbitrary. These are the four integration constants from solving the two second order differential equations. However, the frequency with which the particle goes around the circle is fixed, and given by

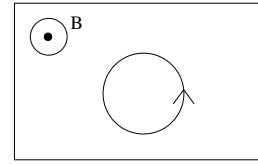


Figure 2:

$$\omega_B = \frac{eB}{m} \quad (1.3)$$

This is called the *cyclotron frequency*.

1.2.2 The Drude Model

Let's now repeat this calculation with two further ingredients. The first is an electric field, \mathbf{E} . This will accelerate the charges and, in the absence of a magnetic field, would result in a current in the direction of \mathbf{E} . The second ingredient is a linear friction term, which is supposed to capture the effect of the electron bouncing off whatever impedes its progress, whether impurities, the underlying lattice or other electrons. The resulting equation of motion is

$$m \frac{d\mathbf{v}}{dt} = -e\mathbf{E} - e\mathbf{v} \times \mathbf{B} - \frac{m\mathbf{v}}{\tau} \quad (1.4)$$

The coefficient τ in the friction term is called the *scattering time*. It can be thought of as the average time between collisions.

The equation of motion (1.4) is the simplest model of charge transport, treating the mobile electrons as if they were classical billiard balls. It is called the *Drude model* and we met it already in the lectures on [Electromagnetism](#).

We're interested in equilibrium solutions of (1.4) which have $d\mathbf{v}/dt = 0$. The velocity of the particle must then solve

$$\mathbf{v} + \frac{e\tau}{m}\mathbf{v} \times \mathbf{B} = -\frac{e\tau}{m}\mathbf{E} \quad (1.5)$$

The current density \mathbf{J} is related to the velocity by

$$\mathbf{J} = -ne\mathbf{v}$$

where n is the density of charge carriers. In matrix notation, (1.5) then becomes

$$\begin{pmatrix} 1 & \omega_B\tau \\ -\omega_B\tau & 1 \end{pmatrix} \mathbf{J} = \frac{e^2n\tau}{m}\mathbf{E}$$

We can invert this matrix to get an equation of the form

$$\mathbf{J} = \sigma\mathbf{E}$$

This equation is known as *Ohm's law*: it tells us how the current flows in response to an electric field. The proportionality constant σ is the *conductivity*. The slight novelty is that, in the presence of a magnetic field, σ is not a single number: it is a matrix. It is sometimes called the *conductivity tensor*. We write it as

$$\sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ -\sigma_{xy} & \sigma_{xx} \end{pmatrix} \quad (1.6)$$

The structure of the matrix, with identical diagonal components, and equal but opposite off-diagonal components, follows from rotational invariance. From the Drude model, we get the explicit expression for the conductivity,

$$\sigma = \frac{\sigma_{DC}}{1 + \omega_B^2\tau^2} \begin{pmatrix} 1 & -\omega_B\tau \\ \omega_B\tau & 1 \end{pmatrix} \quad \text{with} \quad \sigma_{DC} = \frac{ne^2\tau}{m}$$

Here σ_{DC} is the DC conductivity in the absence of a magnetic field. (This is the same result that we derived in the [Electromagnetism](#) lectures). The off-diagonal terms in the matrix are responsible for the Hall effect: in equilibrium, a current in the x -direction requires an electric field with a component in the y -direction.

Although it's not directly relevant for our story, it's worth pausing to think about how we actually approach equilibrium in the Hall effect. We start by putting an electric field in the x -direction. This gives rise to a current density J_x , but this current is deflected due to the magnetic field and bends towards the y -direction. In a finite material, this results in a build up of charge along the edge and an associated electric field E_y . This continues until the electric field E_y cancels the bending of due to the magnetic field, and the electrons then travel only in the x -direction. It's this induced electric field E_y which is responsible for the Hall voltage V_H .

Resistivity vs Resistance

The *resistivity* is defined as the inverse of the conductivity. This remains true when both are matrices,

$$\rho = \sigma^{-1} = \begin{pmatrix} \rho_{xx} & \rho_{xy} \\ -\rho_{xy} & \rho_{yy} \end{pmatrix} \quad (1.7)$$

From the Drude model, we have

$$\rho = \frac{1}{\sigma_{DC}} \begin{pmatrix} 1 & \omega_B \tau \\ -\omega_B \tau & 1 \end{pmatrix} \quad (1.8)$$

The off-diagonal components of the resistivity tensor, $\rho_{xy} = \omega_B \tau / \sigma_{DC}$, have a couple of rather nice properties. First, they are independent of the scattering time τ . This means that they capture something fundamental about the material itself as opposed to the dirty messy stuff that's responsible for scattering.

The second nice property is to do with what we measure. Usually we measure the *resistance* R , which differs from the resistivity ρ by geometric factors. However, for ρ_{xy} , these two things coincide. To see this, consider a sample of material of length L in the y -direction. We drop a voltage V_y in the y -direction and measure the resulting current I_x in the x -direction. The transverse resistance is

$$R_{xy} = \frac{V_y}{I_x} = \frac{L E_y}{L J_x} = \frac{E_y}{J_x} = -\rho_{xy}$$

This has the happy consequence that what we calculate, ρ_{xy} , and what we measure, R_{xy} , are, in this case, the same. In contrast, if we measure the longitudinal resistance R_{xx} then we'll have to divide by the appropriate lengths to extract the resistivity ρ_{xx} . Of course, these lectures are about as theoretical as they come. We're not actually going to measure anything. Just pretend.

While we're throwing different definitions around, here's one more. For a current I_x flowing in the x -direction, and the associated electric field E_y in the y -direction, the *Hall coefficient* is defined by

$$R_H = -\frac{E_y}{J_x B} = \frac{\rho_{xy}}{B}$$

So in the Drude model, we have

$$R_H = \frac{\omega_B}{B\sigma_{DC}} = \frac{1}{ne}$$

As promised, we see that the Hall coefficient depends only on microscopic information about the material: the charge and density of the conducting particles. The Hall coefficient does not depend on the scattering time τ ; it is insensitive to whatever friction processes are at play in the material.

We now have all we need to make an experimental prediction! The two resistivities should be

$$\rho_{xx} = \frac{m}{ne^2\tau} \quad \text{and} \quad \rho_{xy} = \frac{B}{ne}$$

Note that only ρ_{xx} depends on the scattering time τ , and $\rho_{xx} \rightarrow 0$ as scattering processes become less important and $\tau \rightarrow \infty$. If we plot the two resistivities as a function of the magnetic field, then our classical expectation is that they should look the figure on the right.

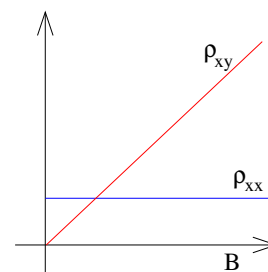


Figure 3:

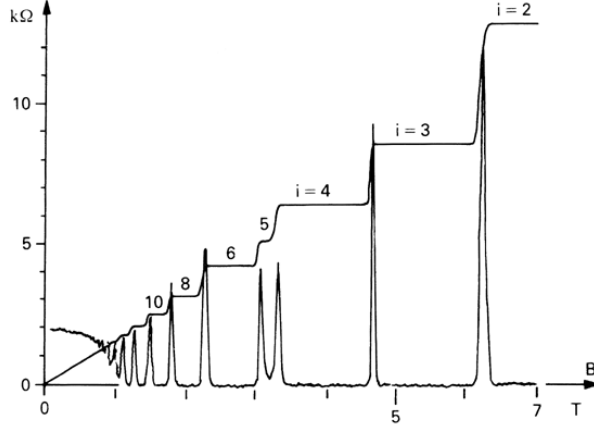
1.3 Quantum Hall Effects

Now we understand the classical expectation. And, of course, this expectation is borne out whenever we can trust classical mechanics. But the world is governed by quantum mechanics. This becomes important at low temperatures and strong magnetic fields where more interesting things can happen.

It's useful to distinguish between two different quantum Hall effects which are associated to two related phenomena. These are called the *integer* and *fractional* quantum Hall effects. Both were first discovered experimentally and only subsequently understood theoretically. Here we summarise the basic facts about these effects. The goal of these lectures is to understand in more detail what's going on.

1.3.1 Integer Quantum Hall Effect

The first experiments exploring the quantum regime of the Hall effect were performed in 1980 by von Klitzing, using samples prepared by Dorda and Pepper¹. The resistivities look like this:



This is the *integer quantum Hall effect*. For this, von Klitzing was awarded the 1985 Nobel prize.

Both the Hall resistivity ρ_{xy} and the longitudinal resistivity ρ_{xx} exhibit interesting behaviour. Perhaps the most striking feature in the data is the fact that the Hall resistivity ρ_{xy} sits on a plateau for a range of magnetic field, before jumping suddenly to the next plateau. On these plateau, the resistivity takes the value

$$\rho_{xy} = \frac{2\pi\hbar}{e^2} \frac{1}{\nu} \quad \nu \in \mathbf{Z} \quad (1.9)$$

The value of ν is measured to be an integer to an extraordinary accuracy. The quantity $2\pi\hbar/e^2$ is called the *quantum of resistivity* (with $-e$, the electron charge). It is now used as the standard for measuring of resistivity. Because ν is measured to be an integer to such remarkable precision – different devices differ only by 3 parts in 10^{10} – the integer quantum Hall effect is now used as the basis for measuring the ratio of fundamental constants $2\pi\hbar/e^2$ sometimes referred to as the von Klitzing constant². This means that, by definition, the $\nu = 1$ state in (1.9) is exactly integer!

¹K. v Klitzing, G. Dorda, M. Pepper, “New Method for High-Accuracy Determination of the Fine-Structure Constant Based on Quantized Hall Resistance”, *Phys. Rev. Lett.* **45** 494.

²Full details of the different quantum Hall set ups and ways to measure the Hall resistivity can be found in B. Jeckelmann and B. Jeanneret, “The quantum Hall effect as an electrical resistance standard”, *Rep. Prog. Phys.* **64**, 1603-1655 (2001).

The centre of each of these plateaux occurs when the magnetic field takes the value

$$B = \frac{2\pi\hbar n}{\nu e} = \frac{n}{\nu} \Phi_0$$

where n is the electron density and $\Phi_0 = 2\pi\hbar/e$ is known as the *flux quantum*. As we will review in Section 2, these are the values of the magnetic field at which the first $\nu \in \mathbf{Z}$ Landau levels are filled. In fact, as we will see, it is very easy to argue that the Hall resistivity should take value (1.9) when ν Landau levels are filled. The surprise is that the plateau exists, with the quantisation persisting over a range of magnetic fields.

There is a clue in the experimental data about the origin of the plateaux. Experimental systems are typically dirty, filled with impurities. The technical name for this is *disorder*. Usually one wants to remove this dirt to get at the underlying physics. Yet, in the quantum Hall effect, as you increase the amount of disorder (within reason) the plateaux become more prominent, not less. In fact, in the absence of disorder, the plateaux are expected to vanish completely. That sounds odd: how can the presence of dirt give rise to something as exact and pure as an integer? This is something we will explain in Section 2.

The longitudinal resistivity ρ_{xx} also exhibits a surprise. When ρ_{xy} sits on a plateau, the longitudinal resistivity vanishes: $\rho_{xx} = 0$. It spikes only when ρ_{xy} jumps to the next plateau.

Usually we would think of a system with $\rho_{xx} = 0$ as a perfect conductor. But there's something a little counter-intuitive about vanishing resistivity in the presence of a magnetic field. To see this, we can return to the simple definition (1.7) which, in components, reads

$$\sigma_{xx} = \frac{\rho_{xx}}{\rho_{xx}^2 + \rho_{xy}^2} \quad \text{and} \quad \sigma_{xy} = \frac{-\rho_{xy}}{\rho_{xx}^2 + \rho_{xy}^2} \quad (1.10)$$

If $\rho_{xy} = 0$ then we get the familiar relation between conductivity and resistivity: $\sigma_{xx} = 1/\rho_{xx}$. But if $\rho_{xy} \neq 0$, then we have the more interesting relation above. In particular, we see

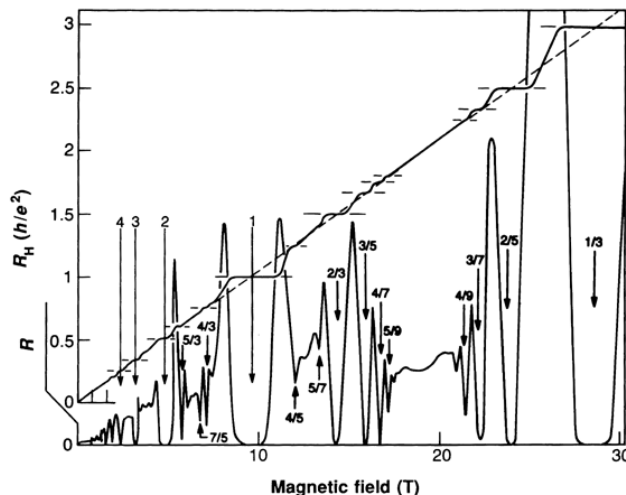
$$\rho_{xx} = 0 \quad \Rightarrow \quad \sigma_{xx} = 0 \quad (\text{if } \rho_{xy} \neq 0)$$

While we would usually call a system with $\rho_{xx} = 0$ a perfect conductor, we would usually call a system with $\sigma_{xx} = 0$ a perfect insulator! What's going on?

This particular surprise has more to do with the words we use to describe the phenomena than the underlying physics. In particular, it has nothing to do with quantum mechanics: this behaviour occurs in the Drude model in the limit $\tau \rightarrow \infty$ where there is no scattering. In this situation, the current is flowing perpendicular to the applied electric field, so $\mathbf{E} \cdot \mathbf{J} = 0$. But recall that $\mathbf{E} \cdot \mathbf{J}$ has the interpretation as the work done in accelerating charges. The fact that this vanishes means that we have a steady current flowing without doing any work and, correspondingly, without any dissipation. The fact that $\sigma_{xx} = 0$ is telling us that no current is flowing in the longitudinal direction (like an insulator) while the fact that $\rho_{xx} = 0$ is telling us that there is no dissipation of energy (like in a perfect conductor).

1.3.2 Fractional Quantum Hall Effect

As the disorder is decreased, the integer Hall plateaux become less prominent. But other plateaux emerge at fractional values. This was discovered in 1982 by Tsui and Störmer using samples prepared by Gossard³. The resistivities look like this:



This is the *fractional quantum Hall effect*. On the plateaux, the Hall resistivity again takes the simple form (1.9), but now with ν a rational number

$$\nu \in \mathbb{Q}$$

Not all fractions appear. The most prominent plateaux sit at $\nu = 1/3, 1/5$ (not shown above) and $2/5$ but there are many more. The vast majority of these have denominators which are odd. But there are exceptions: in particular a clear plateaux has been observed at $\nu = 5/2$.

³D. C. Tsui, H. L. Stormer, and A. C. Gossard, “Two-Dimensional Magnetotransport in the Extreme Quantum Limit”, *Phys. Rev. Lett.* **48** (1982)1559.

As the disorder is decreased, more and more plateaux emerge. Naively, it seems plausible that, in the limit of a perfectly clean sample, we would get an infinite number of plateaux which brings us back to the classical picture of a straight line for ρ_{xy} ! In fact there are arguments that more subtle physics kicks in, and the plateaux persist even in the absence of disorder; we will sketch these arguments in Section 3.3.1.

The integer quantum Hall effect can be understood using free electrons. In contrast, to explain the fractional quantum Hall effect we need to take interactions between electrons into account. This makes the problem much harder and much richer. The basics of the theory were first laid down by Laughlin⁴, but the subject has since expanded in a myriad of different directions. The 1998 Nobel prize was awarded to Tsui, Störmer and Laughlin. Sections 3 onwards will be devoted to aspects of the fractional quantum Hall effect.

Materials

These lectures are unabashedly theoretical. We'll have nothing to say about how one actually constructs these phases of matter in the lab. Here I want to merely throw out a few technical words in an attempt to breed familiarity.

The integer quantum Hall effect was originally discovered in a *Si* MOSFET (this stands for “metal-oxide-semiconductor field-effect transistor”). This is a metal-insulator-semiconductor sandwich, with electrons trapped in the “inversion band” of width $\sim 30\text{\AA}$ between the insulator and semi-conductor. Meanwhile the fractional quantum Hall effect was discovered in a *GaAs-GaAlAs* heterostructure. A lot of the subsequent work was done on this system, and it usually goes by the name *GaAs* (Gallium Arsenide if your chemistry is rusty). In both these systems, the density of electrons is around $n \sim 10^{11} - 10^{12} \text{ cm}^{-2}$.

More recently, both quantum Hall effects have been discovered in graphene, which is a two dimensional material with relativistic electrons. The physics here is similar in spirit, but differs in details.

1.4 Landau Levels

It won't come as a surprise to learn that the physics of the quantum Hall effect involves quantum mechanics. In this section, we will review the quantum mechanics of free particles moving in a background magnetic field and the resulting phenomenon of Landau levels. We will look at these Landau levels in a number of different ways. Each

⁴R. B. Laughlin, “*The Anomalous Quantum Hall Effect: An Incompressible Quantum Fluid with Fractionally Charged Excitations*,” *Phys. Rev. Lett.* **50**, 1395 (1983).

is useful to highlight different aspects of the physics and they will all be important for describing the quantum Hall effects.

Throughout this discussion, we will neglect the spin of the electron. This is more or less appropriate for most physically realised quantum Hall systems. The reason is that in the presence of a magnetic field B there is a Zeeman splitting between the energies of the up and down spins given by $\Delta = 2\mu_B B$ where $\mu_B = e\hbar/2m$ is the Bohr magneton. We will be interested in large magnetic fields where large energies are needed to flip the spin. This means that, if we restrict to low energies, the electrons act as if they are effectively spinless. (We will, however, add a caveat to this argument below.)

Before we get to the quantum theory, we first need to briefly review some of the structure of classical mechanics in the presence of a magnetic field. The Lagrangian for a particle of charge $-e$ and mass m moving in a background magnetic field $\mathbf{B} = \nabla \times \mathbf{A}$ is

$$L = \frac{1}{2}m\dot{\mathbf{x}}^2 - e\dot{\mathbf{x}} \cdot \mathbf{A}$$

Under a gauge transformation, $\mathbf{A} \rightarrow \mathbf{A} + \nabla\alpha$, the Lagrangian changes by a total derivative: $L \rightarrow L - e\dot{\alpha}$. This is enough to ensure that the equations of motion (1.1) remain unchanged under a gauge transformation.

The *canonical momentum* arising from this Lagrangian is

$$\mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{x}}} = m\dot{\mathbf{x}} - e\mathbf{A}$$

This differs from what we called momentum when we were in high school, namely $m\dot{\mathbf{x}}$. We will refer to $m\dot{\mathbf{x}}$ as the *mechanical momentum*.

We can compute the Hamiltonian

$$H = \dot{\mathbf{x}} \cdot \mathbf{p} - L = \frac{1}{2m}(\mathbf{p} + e\mathbf{A})^2$$

If we write the Hamiltonian in terms of the mechanical momentum then it looks the same as it would in the absence of a magnetic field: $H = \frac{1}{2}m\dot{\mathbf{x}}^2$. This is the statement that a magnetic field does no work and so doesn't change the energy of the system. However, there's more to the Hamiltonian framework than just the value of H . We need to remember which variables are canonical. This information is encoded in the Poisson bracket structure of the theory (or, in fancy language, the symplectic structure

on phase space) and, in the quantum theory, is transferred onto commutation relations between operators. The fact that \mathbf{x} and \mathbf{p} are canonical means that

$$\{x_i, p_j\} = \delta_{ij} \quad \text{with} \quad \{x_i, x_j\} = \{p_i, p_j\} = 0 \quad (1.11)$$

Importantly, \mathbf{p} is not gauge invariant. This means that the numerical value of \mathbf{p} can't have any physical meaning since it depends on our choice of gauge. In contrast, the mechanical momentum $m\dot{\mathbf{x}}$ is gauge invariant; it measures what you would physically call “momentum”. But it doesn't have canonical Poisson structure. Specifically, the Poisson bracket of the mechanical momentum with itself is non-vanishing,

$$\{m\dot{x}_i, m\dot{x}_j\} = \{p_i + eA_i, p_j + eA_j\} = -e \left(\frac{\partial A_j}{\partial x^i} - \frac{\partial A_i}{\partial x^j} \right) = -e\epsilon_{ijk}B_k \quad (1.12)$$

Quantisation

Our task is to solve for the spectrum and wavefunctions of the quantum Hamiltonian,

$$H = \frac{1}{2m}(\mathbf{p} + e\mathbf{A})^2 \quad (1.13)$$

Note that we're not going to put hats on operators in this course; you'll just have to remember that they're quantum operators. Since the particle is restricted to lie in the plane, we write $\mathbf{x} = (x, y)$. Meanwhile, we take the magnetic field to be constant and perpendicular to this plane, $\nabla \times \mathbf{A} = B\hat{\mathbf{z}}$. The canonical commutation relations that follow from (1.11) are

$$[x_i, p_j] = i\hbar\delta_{ij} \quad \text{with} \quad [x_i, x_j] = [p_i, p_j] = 0$$

We will first derive the energy spectrum using a purely algebraic method. This is very similar to the algebraic solution of the harmonic oscillator and has the advantage that we don't need to specify a choice of gauge potential \mathbf{A} . The disadvantage is that we don't get to write down specific wavefunctions in terms of the positions of the electrons. We will rectify this in Sections 1.4.1 and 1.4.3.

To proceed, we work with the commutation relations for the mechanical momentum. We'll give it a new name (because the time derivative in $\dot{\mathbf{x}}$ suggests that we're working in the Heisenberg picture which is not necessarily true). We write

$$\boldsymbol{\pi} = \mathbf{p} + e\mathbf{A} = m\dot{\mathbf{x}} \quad (1.14)$$

Then the commutation relations following from the Poisson bracket (1.12) are

$$[\pi_x, \pi_y] = -ie\hbar B \quad (1.15)$$

At this point we introduce new variables. These are raising and lowering operators, entirely analogous to those that we use in the harmonic oscillator. They are defined by

$$a = \frac{1}{\sqrt{2e\hbar B}}(\pi_x - i\pi_y) \quad \text{and} \quad a^\dagger = \frac{1}{\sqrt{2e\hbar B}}(\pi_x + i\pi_y)$$

The commutation relations for $\boldsymbol{\pi}$ then tell us that a and a^\dagger obey

$$[a, a^\dagger] = 1$$

which are precisely the commutation relations obeyed by the raising and lowering operators of the harmonic oscillator. Written in terms of these operators, the Hamiltonian (1.13) even takes the same form as that of the harmonic oscillator

$$H = \frac{1}{2m}\boldsymbol{\pi} \cdot \boldsymbol{\pi} = \hbar\omega_B \left(a^\dagger a + \frac{1}{2} \right)$$

where $\omega_B = eB/m$ is the cyclotron frequency that we met previously (1.3).

Now it's simple to finish things off. We can construct the Hilbert space in the same way as the harmonic oscillator: we first introduce a ground state $|0\rangle$ obeying $a|0\rangle = 0$ and build the rest of the Hilbert space by acting with a^\dagger ,

$$a^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle \quad \text{and} \quad a|n\rangle = \sqrt{n}|n-1\rangle$$

The state $|n\rangle$ has energy

$$E_n = \hbar\omega_B \left(n + \frac{1}{2} \right) \quad n \in \mathbf{N} \quad (1.16)$$

We learn that in the presence of a magnetic field, the energy levels of a particle become equally spaced, with the gap between each level proportional to the magnetic field B . The energy levels are called *Landau levels*. Notice that this is not a small change: the spectrum looks very very different from that of a free particle in the absence of a magnetic field.

There's something a little disconcerting about the above calculation. We started with a particle moving in a plane. This has two degrees of freedom. But we ended up writing this in terms of the harmonic oscillator which has just a single degree of freedom. It seems like we lost something along the way! And, in fact, we did. The energy levels (1.16) are the correct spectrum of the theory but, unlike for the harmonic oscillator, it turns out that each energy level does not have a unique state associated to it. Instead there is a degeneracy of states. A wild degeneracy. We will return to the algebraic approach in Section 1.4.3 and demonstrate this degeneracy. But it's simplest to first turn to a specific choice of the gauge potential \mathbf{A} , which we do shortly.

A Quick Aside: The role of spin

The splitting between Landau levels is $\Delta = \hbar\omega_B = e\hbar B/m$. But, for free electrons, this precisely coincides with the Zeeman splitting $\Delta = g\mu_B B$ between spins, where $\mu_B = e\hbar/2m$ is the Bohr magneton and, famously, $g = 2$. It looks as if the spin up particles in Landau level n have exactly the same energy as the spin down particles in level $n + 1$. In fact, in real materials, this does not happen. The reason is twofold. First, the true value of the cyclotron frequency is $\omega_B = eB/m_{\text{eff}}$, where m_{eff} is the effective mass of the electron moving in its environment. Second, the g factor can also vary due to effects of band structure. For *GaAs*, the result is that the Zeeman energy is typically about 70 times smaller than the cyclotron energy. This means that first the $n = 0$ spin-up Landau level fills, then the $n = 0$ spin-down, then the $n = 1$ spin-up and so on. For other materials (such as the interface between *ZnO* and *MnZnO*) the relative size of the energies can be flipped and you can fill levels in a different order. This results in different fractional quantum Hall states. In these notes, we will mostly ignore these issues to do with spin⁵. (One exception is Section 3.3.4 where we discuss wavefunctions for particles with spin).

1.4.1 Landau Gauge

To find wavefunctions corresponding to the energy eigenstates, we first need to specify a gauge potential \mathbf{A} such that

$$\nabla \times \mathbf{A} = B\hat{\mathbf{z}}$$

There is, of course, not a unique choice. In this section and the next we will describe two different choices of \mathbf{A} .

In this section, we work with the choice

$$\mathbf{A} = xB\hat{\mathbf{y}} \tag{1.17}$$

This is called *Landau gauge*. Note that the magnetic field \mathbf{B} is invariant under both translational symmetry and rotational symmetry in the (x, y) -plane. However, the choice of \mathbf{A} is not; it breaks translational symmetry in the x direction (but not in

⁵It is far from clear that it is acceptable to ignore these issues! For example, in these lectures, we will treat the $\nu = 1$ quantum Hall state in the framework of non-interacting electrons. But it seems likely the $\nu = 1$ state in, say, *GaAs* can only be understood by including interactions between spins. This story was initiated in the paper by “*Skyrmions and the crossover from the integer to fractional quantum Hall effect at small Zeeman energies*” by Sondhi, Karlhede, Kivelson, and Rezayi, and originally came with the pithy punchline “ $\nu = 1$ is a fraction too”.

the y direction) and rotational symmetry. This means that, while the physics will be invariant under all symmetries, the intermediate calculations will not be manifestly invariant. This kind of compromise is typical when dealing with magnetic field.

The Hamiltonian (1.13) becomes

$$H = \frac{1}{2m} (p_x^2 + (p_y + eBx)^2)$$

Because we have manifest translational invariance in the y direction, we can look for energy eigenstates which are also eigenstates of p_y . These, of course, are just plane waves in the y direction. This motivates an ansatz using the separation of variables,

$$\psi_k(x, y) = e^{iky} f_k(x) \quad (1.18)$$

Acting on this wavefunction with the Hamiltonian, we see that the operator p_y just gets replaced by its eigenvalue $\hbar k$,

$$H\psi_k(x, y) = \frac{1}{2m} (p_x^2 + (\hbar k + eBx)^2) \psi_k(x, y) \equiv H_k \psi_k(x, y)$$

But this is now something very familiar: it's the Hamiltonian for a harmonic oscillator in the x direction, with the centre displaced from the origin,

$$H_k = \frac{1}{2m} p_x^2 + \frac{m\omega_B^2}{2} (x + kl_B^2)^2 \quad (1.19)$$

The frequency of the harmonic oscillator is again the cyclotron frequency $\omega_B = eB/m$, and we've also introduced a length scale l_B . This is a characteristic length scale which governs any quantum phenomena in a magnetic field. It is called the *magnetic length*.

$$l_B = \sqrt{\frac{\hbar}{eB}}$$

To give you some sense for this, in a magnetic field of $B = 1$ *Tesla*, the magnetic length for an electron is $l_B \approx 2.5 \times 10^{-8}$ *m*.

Something rather strange has happened in the Hamiltonian (1.19): the momentum in the y direction, $\hbar k$, has turned into the position of the harmonic oscillator in the x direction, which is now centred at $x = -kl_B^2$.

Just as in the algebraic approach above, we've reduced the problem to that of the harmonic oscillator. The energy eigenvalues are

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

But now we can also write down the explicit wavefunctions. They depend on two quantum numbers, $n \in \mathbf{N}$ and $k \in \mathbf{R}$,

$$\psi_{n,k}(x, y) \sim e^{iky} H_n(x + kl_B^2) e^{-(x+kl_B^2)^2/2l_B^2} \quad (1.20)$$

with H_n the usual Hermite polynomial wavefunctions of the harmonic oscillator. The \sim reflects the fact that we have made no attempt to normalise these wavefunctions.

The wavefunctions look like strips, extended in the y direction but exponentially localised around $x = -kl_B^2$ in the x direction. However, the large degeneracy means that by taking linear combinations of these states, we can cook up wavefunctions that have pretty much any shape you like. Indeed, in the next section we will choose a different \mathbf{A} and see very different profiles for the wavefunctions.

Degeneracy

One advantage of this approach is that we can immediately see the degeneracy in each Landau level. The wavefunction (1.20) depends on two quantum numbers, n and k but the energy levels depend only on n . Let's now see how large this degeneracy is.

To do this, we need to restrict ourselves to a finite region of the (x, y) -plane. We pick a rectangle with sides of lengths L_x and L_y . We want to know how many states fit inside this rectangle.

Having a finite size L_y is like putting the system in a box in the y -direction. We know that the effect of this is to quantise the momentum k in units of $2\pi/L_y$.

Having a finite size L_x is somewhat more subtle. The reason is that, as we mentioned above, the gauge choice (1.17) does not have manifest translational invariance in the x -direction. This means that our argument will be a little heuristic. Because the wavefunctions (1.20) are exponentially localised around $x = -kl_B^2$, for a finite sample restricted to $0 \leq x \leq L_x$ we would expect the allowed k values to range between $-L_x/l_B^2 \leq k \leq 0$. The end result is that the number of states is

$$\mathcal{N} = \frac{L_y}{2\pi} \int_{-L_x/l_B^2}^0 dk = \frac{L_x L_y}{2\pi l_B^2} = \frac{eBA}{2\pi\hbar} \quad (1.21)$$

where $A = L_x L_y$ is the area of the sample. Despite the slight approximation used above, this turns out to be the exact answer for the number of states on a torus. (One can do better taking the wavefunctions on a torus to be elliptic theta functions).

The degeneracy (1.21) is very very large. There are a macroscopic number of states in each Landau level. The resulting spectrum looks like the figure on the right, with $n \in \mathbf{N}$ labelling the Landau levels and the energy independent of k . This degeneracy will be responsible for much of the interesting physics of the fractional quantum Hall effect that we will meet in Section 3.

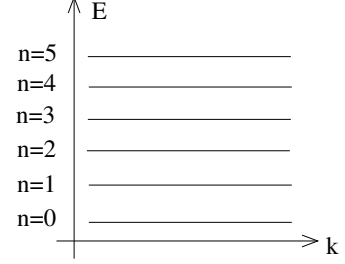


Figure 4: Landau Levels

It is common to introduce some new notation to describe the degeneracy (1.21). We write

$$\mathcal{N} = \frac{AB}{\Phi_0} \quad \text{with} \quad \Phi_0 = \frac{2\pi\hbar}{e} \quad (1.22)$$

Φ_0 is called the *quantum of flux*. It can be thought of as the magnetic flux contained within the area $2\pi l_B^2$. It plays an important role in a number of quantum phenomena in the presence of magnetic fields.

1.4.2 Turning on an Electric Field

The Landau gauge is useful for working in rectangular geometries. One of the things that is particularly easy in this gauge is the addition of an electric field E in the x direction. We can implement this by the addition of an electric potential $\phi = -Ex$. The resulting Hamiltonian is

$$H = \frac{1}{2m} (p_x^2 + (p_y + eBx)^2) + eEx \quad (1.23)$$

We can again use the ansatz (1.18). We simply have to complete the square to again write the Hamiltonian as that of a displaced harmonic oscillator. The states are related to those that we had previously, but with a shifted argument

$$\psi(x, y) = \psi_{n,k}(x + mE/eB^2, y) \quad (1.24)$$

and the energies are now given by

$$E_{n,k} = \hbar\omega_B \left(n + \frac{1}{2} \right) - eE \left(kl_B^2 + \frac{eE}{m\omega_B^2} \right) + \frac{m}{2} \frac{E^2}{B^2} \quad (1.25)$$

This is interesting. The degeneracy in each Landau level has now been lifted. The energy in each level now depends linearly on k , as shown in the figure.

Because the energy now depends on the momentum, it means that states now drift in the y direction. The group velocity is

$$v_y = \frac{1}{\hbar} \frac{\partial E_{n,k}}{\partial k} = -\frac{e}{\hbar} E l_B^2 = -\frac{E}{B} \quad (1.26)$$

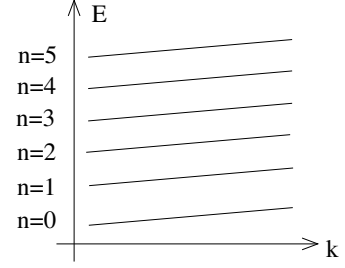


Figure 5: Landau Levels in an electric field

This result is one of the surprising joys of classical physics:

if you put an electric field \mathbf{E} perpendicular to a magnetic field \mathbf{B} then the cyclotron orbits of the electron drift. But they don't drift in the direction of the electric field! Instead they drift in the direction $\mathbf{E} \times \mathbf{B}$. Here we see the quantum version of this statement.

The fact that the particles are now moving also provides a natural interpretation of the energy (1.25). A wavepacket with momentum k is now localised at position $x = -kl_B^2 - eE/m\omega_B^2$; the middle term above can be thought of as the potential energy of this wavepacket. The final term can be thought of as the kinetic energy for the particle in the y direction: $\frac{1}{2}mv_y^2$.

1.4.3 Symmetric Gauge

Having understood the basics of Landau levels, we're now going to do it all again. This time we'll work in *symmetric gauge*, with

$$\mathbf{A} = -\frac{1}{2}\mathbf{r} \times \mathbf{B} = -\frac{yB}{2}\hat{\mathbf{x}} + \frac{xB}{2}\hat{\mathbf{y}} \quad (1.27)$$

This choice of gauge breaks translational symmetry in both the x and the y directions. However, it does preserve rotational symmetry about the origin. This means that angular momentum is a good quantum number.

The main reason for studying Landau levels in symmetric gauge is that this is most convenient language for describing the fractional quantum Hall effect. We shall look at this in Section 3. However, as we now see, there are also a number of pretty things that happen in symmetric gauge.

The Algebraic Approach Revisited

At the beginning of this section, we provided a simple algebraic derivation of the energy spectrum (1.16) of a particle in a magnetic field. But we didn't provide an algebraic

derivation of the degeneracies of these Landau levels. Here we rectify this. As we will see, this derivation only really works in the symmetric gauge.

Recall that the algebraic approach uses the mechanical momenta $\boldsymbol{\pi} = \mathbf{p} + e\mathbf{A}$. This is gauge invariant, but non-canonical. We can use this to build ladder operators $a = (\pi_x - i\pi_y)/\sqrt{2e\hbar B}$ which obey $[a, a^\dagger] = 1$. In terms of these creation operators, the Hamiltonian takes the harmonic oscillator form,

$$H = \frac{1}{2m}\boldsymbol{\pi} \cdot \boldsymbol{\pi} = \hbar\omega_B \left(a^\dagger a + \frac{1}{2} \right)$$

To see the degeneracy in this language, we start by introducing yet another kind of “momentum”,

$$\tilde{\boldsymbol{\pi}} = \mathbf{p} - e\mathbf{A} \tag{1.28}$$

This differs from the mechanical momentum (1.14) by the minus sign. This means that, in contrast to $\boldsymbol{\pi}$, this new momentum is *not* gauge invariant. We should be careful when interpreting the value of $\tilde{\boldsymbol{\pi}}$ since it can change depending on choice of gauge potential \mathbf{A} .

The commutators of this new momenta differ from (1.15) only by a minus sign

$$[\tilde{\pi}_x, \tilde{\pi}_y] = ie\hbar B \tag{1.29}$$

However, the lack of gauge invariance shows up when we take the commutators of $\boldsymbol{\pi}$ and $\tilde{\boldsymbol{\pi}}$. We find

$$[\pi_x, \tilde{\pi}_x] = 2ie\hbar \frac{\partial A_x}{\partial x} \quad , \quad [\pi_y, \tilde{\pi}_y] = 2ie\hbar \frac{\partial A_y}{\partial y} \quad , \quad [\pi_x, \tilde{\pi}_y] = [\pi_y, \tilde{\pi}_x] = ie\hbar \left(\frac{\partial A_x}{\partial y} + \frac{\partial A_y}{\partial x} \right)$$

This is unfortunate. It means that we cannot, in general, simultaneously diagonalise $\tilde{\boldsymbol{\pi}}$ and the Hamiltonian H which, in turn, means that we can't use $\tilde{\pi}$ to tell us about other quantum numbers in the problem.

There is, however, a happy exception to this. In symmetric gauge (1.27) all these commutators vanish and we have

$$[\pi_i, \tilde{\pi}_j] = 0$$

We can now define a second pair of raising and lowering operators,

$$b = \frac{1}{\sqrt{2e\hbar B}} (\tilde{\pi}_x + i\tilde{\pi}_y) \quad \text{and} \quad b^\dagger = \frac{1}{\sqrt{2e\hbar B}} (\tilde{\pi}_x - i\tilde{\pi}_y)$$

These too obey

$$[b, b^\dagger] = 1$$

It is this second pair of creation operators that provide the degeneracy of the Landau levels. We define the ground state $|0, 0\rangle$ to be annihilated by both lowering operators, so that $a|0, 0\rangle = b|0, 0\rangle = 0$. Then the general state in the Hilbert space is $|n, m\rangle$ defined by

$$|n, m\rangle = \frac{a^{\dagger n} b^{\dagger m}}{\sqrt{n!m!}} |0, 0\rangle$$

The energy of this state is given by the usual Landau level expression (1.16); it depends on n but not on m .

The Lowest Landau Level

Let's now construct the wavefunctions in the symmetric gauge. We're going to focus attention on the lowest Landau level, $n = 0$, since this will be of primary interest when we come to discuss the fractional quantum Hall effect. The states in the lowest Landau level are annihilated by a , meaning $a|0, m\rangle = 0$. The trick is to convert this into a differential equation. The lowering operator is

$$\begin{aligned} a &= \frac{1}{\sqrt{2e\hbar B}} (\pi_x - i\pi_y) \\ &= \frac{1}{\sqrt{2e\hbar B}} (p_x - ip_y + e(A_x - iA_y)) \\ &= \frac{1}{\sqrt{2e\hbar B}} \left(-i\hbar \left(\frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right) + \frac{eB}{2} (-y - ix) \right) \end{aligned}$$

At this stage, it's useful to work in complex coordinates on the plane. We introduce

$$z = x - iy \quad \text{and} \quad \bar{z} = x + iy$$

Note that this is the opposite to how we would normally define these variables! It's annoying but it's because we want the wavefunctions below to be holomorphic rather than anti-holomorphic. (An alternative would be to work with magnetic fields $B < 0$ in which case we get to use the usual definition of holomorphic. However, we'll stick with our choice above throughout these lectures). We also introduce the corresponding holomorphic and anti-holomorphic derivatives

$$\partial = \frac{1}{2} \left(\frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right) \quad \text{and} \quad \bar{\partial} = \frac{1}{2} \left(\frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right)$$

which obey $\partial z = \bar{\partial} \bar{z} = 1$ and $\partial \bar{z} = \bar{\partial} z = 0$. In terms of these holomorphic coordinates, a takes the simple form

$$a = -i\sqrt{2} \left(l_B \bar{\partial} + \frac{z}{4l_B} \right)$$

and, correspondingly,

$$a^\dagger = -i\sqrt{2} \left(l_B \partial - \frac{\bar{z}}{4l_B} \right)$$

which we've chosen to write in terms of the magnetic length $l_B = \sqrt{\hbar/eB}$. The lowest Landau level wavefunctions $\psi_{LLL}(z, \bar{z})$ are then those which are annihilated by this differential operator. But this is easily solved: they are

$$\psi_{LLL}(z, \bar{z}) = f(z) e^{-|z|^2/4l_B^2}$$

for any holomorphic function $f(z)$.

We can construct the specific states $|0, m\rangle$ in the lowest Landau level by similarly writing b and b^\dagger as differential operators. We find

$$b = -i\sqrt{2} \left(l_B \partial + \frac{\bar{z}}{4l_B} \right) \quad \text{and} \quad b^\dagger = -i\sqrt{2} \left(l_B \bar{\partial} - \frac{z}{4l_B} \right)$$

The lowest state $\psi_{LLL, m=0}$ is annihilated by both a and b . There is a unique such state given by

$$\psi_{LLL, m=0} \sim e^{-|z|^2/4l_B^2}$$

We can now construct the higher states by acting with b^\dagger . Each time we do this, we pull down a factor of $z/2l_B$. This gives us a basis of lowest Landau level wavefunctions in terms of holomorphic monomials

$$\psi_{LLL, m} \sim \left(\frac{z}{l_B} \right)^m e^{-|z|^2/4l_B^2} \tag{1.30}$$

This particular basis of states has another advantage: these are eigenstates of angular momentum. To see this, we define angular momentum operator,

$$J = i\hbar \left(x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right) = \hbar(z\partial - \bar{z}\bar{\partial}) \tag{1.31}$$

Then, acting on these lowest Landau level states we have

$$J\psi_{LLL, m} = \hbar m \psi_{LLL, m}$$

The wavefunctions (1.30) provide a basis for the lowest Landau level. But it is a simple matter to extend this to write down wavefunctions for all high Landau levels; we simply need to act with the raising operator $a^\dagger = -i\sqrt{2}(l_B\partial - \bar{z}/4l_B)$. However, we won't have any need for the explicit forms of these higher Landau level wavefunctions in what follows.

Degeneracy Revisited

In symmetric gauge, the profiles of the wavefunctions (1.30) form concentric rings around the origin. The higher the angular momentum m , the further out the ring. This, of course, is very different from the strip-like wavefunctions that we saw in Landau gauge (1.20). You shouldn't read too much into this other than the fact that the profile of the wavefunctions is not telling us anything physical as it is not gauge invariant.

However, it's worth seeing how we can see the degeneracy of states in symmetric gauge. The wavefunction with angular momentum m is peaked on a ring of radius $r = \sqrt{2m}l_B$. This means that in a disc shaped region of area $A = \pi R^2$, the number of states is roughly (the integer part of)

$$\mathcal{N} = R^2/2l_B^2 = A/2\pi l_B^2 = \frac{eBA}{2\pi\hbar}$$

which agrees with our earlier result (1.21).

There is yet another way of seeing this degeneracy that makes contact with the classical physics. In Section 1.2, we reviewed the classical motion of particles in a magnetic field. They go in circles. The most general solution to the classical equations of motion is given by (1.2),

$$x(t) = X - R \sin(\omega_B t + \phi) \quad \text{and} \quad y(t) = Y + R \cos(\omega_B t + \phi) \quad (1.32)$$

Let's try to tally this with our understanding of the exact quantum states in terms of Landau levels. To do this, we'll think about the coordinates labelling the centre of the orbit (X, Y) as quantum operators. We can rearrange (1.32) to give

$$\begin{aligned} X &= x(t) + R \sin(\omega_B t + \phi) = x - \frac{\dot{y}}{\omega_B} = x - \frac{\pi_y}{m\omega_B} \\ Y &= y(t) - R \cos(\omega_B t + \phi) = y + \frac{\dot{x}}{\omega_B} = y + \frac{\pi_x}{m\omega_B} \end{aligned} \quad (1.33)$$

This feels like something of a slight of hand, but the end result is what we wanted: we have the centre of mass coordinates in terms of familiar quantum operators. Indeed, one can check that under time evolution, we have

$$i\hbar\dot{X} = [X, H] = 0 \quad , \quad i\hbar\dot{Y} = [Y, H] = 0 \quad (1.34)$$

confirming the fact that these are constants of motion.

The definition of the centre of the orbit (X, Y) given above holds in any gauge. If we now return to symmetric gauge we can replace the x and y coordinates appearing here with the gauge potential (1.27). We end up with

$$X = \frac{1}{eB} (2eA_y - \pi_y) = -\frac{\tilde{\pi}_y}{eB} \quad \text{and} \quad Y = \frac{1}{eB} (-2eA_x + \pi_x) = \frac{\tilde{\pi}_x}{eB}$$

where, finally, we've used the expression (1.28) for the “alternative momentum” $\tilde{\pi}$. We see that, in symmetric gauge, the alternative momentum has the nice interpretation of the centre of the orbit! The commutation relation (1.29) then tells us that the positions of the orbit in the (X, Y) plane fail to commute with each other,

$$[X, Y] = i l_B^2 \tag{1.35}$$

The lack of commutivity is precisely the magnetic length $l_B^2 = \hbar/eB$. The Heisenberg uncertainty principle now means that we can't localise states in both the X coordinate and the Y coordinate: we have to find a compromise. In general, the uncertainty is given by

$$\Delta X \Delta Y = \frac{l_B^2}{2}$$

A naive (Bohr-Sommerfeld) semi-classical count of the states then comes from taking the plane and parcelling it up into regions of area $2\pi l_B^2$. The number of states in an area A is then

$$\mathcal{N} = \frac{A}{\Delta X \Delta Y} = \frac{A}{2\pi l_B^2} = \frac{eBA}{2\pi\hbar}$$

which is the counting that we've already seen above.

1.5 Berry Phase

There is one last topic that we need to review before we can start the story of the quantum Hall effect. This is the subject of Berry phase or, more precisely, the Berry holonomy⁶. This is not a topic which is relevant just in quantum Hall physics: it has applications in many areas of quantum mechanics and will arise over and over again in different guises in these lectures. Moreover, it is a topic which perhaps captures the spirit of the quantum Hall effect better than any other, for the Berry phase is the simplest demonstration of how geometry and topology can emerge from quantum mechanics. As we will see in these lectures, this is the heart of the quantum Hall effect.

⁶An excellent review of this subject can be found in the book *Geometric Phases in Physics* by Wilczek and Shapere

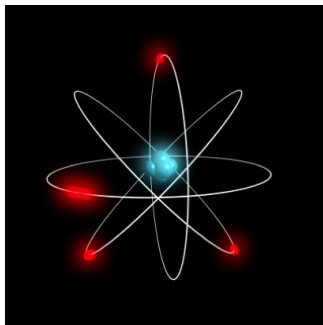


Figure 6: The degrees of freedom x .



Figure 7: The parameters λ .

1.5.1 Abelian Berry Phase and Berry Connection

We'll describe the Berry phase arising for a general Hamiltonian which we write as

$$H(x^a; \lambda^i)$$

As we've illustrated, the Hamiltonian depends on two different kinds of variables. The x^a are the degrees of freedom of the system. These are the things that evolve dynamically, the things that we want to solve for in any problem. They are typically things like the positions or spins of particles.

In contrast, the other variables λ^i are the parameters of the Hamiltonian. They are fixed, with their values determined by some external apparatus that probably involves knobs and dials and flashing lights and things as shown above. We don't usually exhibit the dependence of H on these variables⁷.

Here's the game. We pick some values for the parameters λ and place the system in a specific energy eigenstate $|\psi\rangle$ which, for simplicity, we will take to be the ground state. We assume this ground state is unique (an assumption which we will later relax in Section 1.5.4). Now we very slowly vary the parameters λ . The Hamiltonian changes so, of course, the ground state also changes; it is $|\psi(\lambda(t))\rangle$.

There is a theorem in quantum mechanics called the *adiabatic theorem*. This states that if we place a system in a non-degenerate energy eigenstate and vary parameters sufficiently slowly, then the system will cling to that energy eigenstate. It won't be excited to any higher or lower states.

⁷One exception is the classical subject of adiabatic invariants, where we also think about how H depends on parameters λ . See section 4.6 of the notes on *Classical Dynamics*.

There is one caveat to the adiabatic theorem. How slow you have to be in changing the parameters depends on the energy gap from the state you're in to the nearest other state. This means that if you get *level crossing*, where another state becomes degenerate with the one you're in, then all bets are off. When the states separate again, there's no simple way to tell which linear combinations of the state you now sit in. However, level crossings are rare in quantum mechanics. In general, you have to tune three parameters to specific values in order to get two states to have the same energy. This follows by thinking about the a general Hermitian 2×2 matrix which can be viewed as the Hamiltonian for the two states of interest. The general Hermitian 2×2 matrix depends on 4 parameters, but its eigenvalues only coincide if it is proportional to the identity matrix. This means that three of those parameters have to be set to zero.

The idea of the Berry phase arises in the following situation: we vary the parameters λ but, ultimately, we put them back to their starting values. This means that we trace out a closed path in the space of parameters. We will assume that this path did not go through a point with level crossing. The question is: what state are we now in?

The adiabatic theorem tells us most of the answer. If we started in the ground state, we also end up in the ground state. The only thing left uncertain is the phase of this new state

$$|\psi\rangle \rightarrow e^{i\gamma}|\psi\rangle$$

We often think of the overall phase of a wavefunction as being unphysical. But that's not the case here because this is a *phase difference*. For example, we could have started with two states and taken only one of them on this journey while leaving the other unchanged. We could then interfere these two states and the phase $e^{i\gamma}$ would have physical consequence.

So what is the phase $e^{i\gamma}$? There are two contributions. The first is simply the dynamical phase $e^{-iEt/\hbar}$ that is there for any energy eigenstate, even if the parameters don't change. But there is also another, less obvious contribution to the phase. This is the *Berry phase*.

Computing the Berry Phase

The wavefunction of the system evolves through the time-dependent Schrödinger equation

$$i\hbar \frac{\partial |\psi\rangle}{\partial t} = H(\lambda(t))|\psi\rangle \quad (1.36)$$

For every choice of the parameters λ , we introduce a ground state with some fixed choice of phase. We call these reference states $|n(\lambda)\rangle$. There is no canonical way to do this; we just make an arbitrary choice. We'll soon see how this choice affects the final answer. The adiabatic theorem means that the ground state $|\psi(t)\rangle$ obeying (1.36) can be written as

$$|\psi(t)\rangle = U(t) |n(\lambda(t))\rangle \quad (1.37)$$

where $U(t)$ is some time dependent phase. If we pick the $|n(\lambda(t=0))\rangle = |\psi(t=0)\rangle$ then we have $U(t=0) = 1$. Our task is then to determine $U(t)$ after we've taken λ around the closed path and back to where we started.

There's always the dynamical contribution to the phase, given by $e^{-i \int dt E_0(t)/\hbar}$ where E_0 is the ground state energy. This is not what's interesting here and we will ignore it simply by setting $E_0(t) = 0$. However, there is an extra contribution. This arises by plugging the adiabatic ansatz into (1.36), and taking the overlap with $\langle\psi|$. We have

$$\langle\psi|\dot{\psi}\rangle = \dot{U}U^* + \langle n|\dot{n}\rangle = 0$$

where we've used the fact that, instantaneously, $H(\lambda)|n(\lambda)\rangle = 0$ to get zero on the right-hand side. (Note: this calculation is actually a little more subtle than it looks. To do a better job we would have to look more closely at corrections to the adiabatic evolution (1.37)). This gives us an expression for the time dependence of the phase U ,

$$U^*\dot{U} = -\langle n|\dot{n}\rangle = -\langle n|\frac{\partial}{\partial\lambda^i}|n\rangle \dot{\lambda}^i \quad (1.38)$$

It is useful to define the *Berry connection*

$$\mathcal{A}_i(\lambda) = -i\langle n|\frac{\partial}{\partial\lambda^i}|n\rangle \quad (1.39)$$

so that (1.38) reads

$$\dot{U} = -i\mathcal{A}_i \dot{\lambda}^i U$$

But this is easily solved. We have

$$U(t) = \exp\left(-i \int \mathcal{A}_i(\lambda) \dot{\lambda}^i dt\right)$$

Our goal is to compute the phase $U(t)$ after we've taken a closed path C in parameter space. This is simply

$$e^{i\gamma} = \exp\left(-i \oint_C \mathcal{A}_i(\lambda) d\lambda^i\right) \quad (1.40)$$

This is the *Berry phase*. Note that it doesn't depend on the time taken to change the parameters. It does, however, depend on the path taken through parameter space.

The Berry Connection

Above we introduced the idea of the Berry connection (1.39). This is an example of a kind of object that you've seen before: it is like the gauge potential in electromagnetism! Let's explore this analogy a little further.

In the relativistic form of electromagnetism, we have a gauge potential $A_\mu(x)$ where $\mu = 0, 1, 2, 3$ and x are coordinates over Minkowski spacetime. There is a redundancy in the description of the gauge potential: all physics remains invariant under the gauge transformation

$$A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu \omega \quad (1.41)$$

for any function $\omega(x)$. In our course on electromagnetism, we were taught that if we want to extract the physical information contained in A_μ , we should compute the field strength

$$F_{\mu\nu} = \frac{\partial A_\mu}{\partial x^\nu} - \frac{\partial A_\nu}{\partial x^\mu}$$

This contains the electric and magnetic fields. It is invariant under gauge transformations.

Now let's compare this to the Berry connection $\mathcal{A}_i(\lambda)$. Of course, this no longer depends on the coordinates of Minkowski space; instead it depends on the parameters λ^i . The number of these parameters is arbitrary; let's suppose that we have d of them. This means that $i = 1, \dots, d$. In the language of differential geometry $\mathcal{A}_i(\lambda)$ is said to be a one-form over the space of parameters, while $A_i(x)$ is said to be a one-form over Minkowski space.

There is also a redundancy in the information contained in the Berry connection $\mathcal{A}_i(\lambda)$. This follows from the arbitrary choice we made in fixing the phase of the reference states $|n(\lambda)\rangle$. We could just as happily have chosen a different set of reference states which differ by a phase. Moreover, we could pick a different phase for every choice of parameters λ ,

$$|n'(\lambda)\rangle = e^{i\omega(\lambda)} |n(\lambda)\rangle$$

for any function $\omega(\lambda)$. If we compute the Berry connection arising from this new choice, we have

$$\mathcal{A}'_i = -i \langle n' | \frac{\partial}{\partial \lambda^i} | n' \rangle = \mathcal{A}_i + \frac{\partial \omega}{\partial \lambda^i} \quad (1.42)$$

This takes the same form as the gauge transformation (1.41).

Following the analogy with electromagnetism, we might expect that the physical information in the Berry connection can be found in the gauge invariant field strength which, mathematically, is known as the *curvature* of the connection,

$$\mathcal{F}_{ij}(\lambda) = \frac{\partial \mathcal{A}_i}{\partial \lambda^j} - \frac{\partial \mathcal{A}_j}{\partial \lambda^i}$$

It's certainly true that \mathcal{F} contains some physical information about our quantum system and we'll have use of this in later sections. But it's not the only gauge invariant quantity of interest. In the present context, the most natural thing to compute is the Berry phase (1.40). Importantly, this too is independent of the arbitrariness arising from the gauge transformation (1.42). This is because $\oint \partial_i \omega d\lambda^i = 0$. In fact, it's possible to write the Berry phase in terms of the field strength using the higher-dimensional version of Stokes' theorem

$$e^{i\gamma} = \exp \left(-i \oint_C \mathcal{A}_i(\lambda) d\lambda^i \right) = \exp \left(-i \int_S \mathcal{F}_{ij} dS^{ij} \right) \quad (1.43)$$

where S is a two-dimensional surface in the parameter space bounded by the path C .

1.5.2 An Example: A Spin in a Magnetic Field

The standard example of the Berry phase is very simple. It is a spin, with a Hilbert space consisting of just two states. The spin is placed in a magnetic field \vec{B} , with Hamiltonian which we take to be

$$H = -\vec{B} \cdot \vec{\sigma} + B$$

with $\vec{\sigma}$ the triplet of Pauli matrices and $B = |\vec{B}|$. The offset ensures that the ground state always has vanishing energy. Indeed, this Hamiltonian has two eigenvalues: 0 and $+2B$. We denote the ground state as $|\downarrow\rangle$ and the excited state as $|\uparrow\rangle$,

$$H|\downarrow\rangle = 0 \quad \text{and} \quad H|\uparrow\rangle = 2B|\uparrow\rangle$$

Note that these two states are non-degenerate as long as $\vec{B} \neq 0$.

We are going to treat the magnetic field as the parameters, so that $\lambda^i \equiv \vec{B}$ in this example. Be warned: this means that things are about to get confusing because we'll be talking about Berry connections \mathcal{A}_i and curvatures \mathcal{F}_{ij} over the space of magnetic fields. (As opposed to electromagnetism where we talk about magnetic fields over actual space).

The specific form of $|\uparrow\rangle$ and $|\downarrow\rangle$ will depend on the orientation of \vec{B} . To provide more explicit forms for these states, we write the magnetic field \vec{B} in spherical polar coordinates

$$\vec{B} = \begin{pmatrix} B \sin \theta \cos \phi \\ B \sin \theta \sin \phi \\ B \cos \theta \end{pmatrix}$$

with $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$ The Hamiltonian then reads

$$H = -B \begin{pmatrix} \cos \theta - 1 & e^{-i\phi} \sin \theta \\ e^{+i\phi} \sin \theta & -\cos \theta - 1 \end{pmatrix}$$

In these coordinates, two normalised eigenstates are given by

$$|\downarrow\rangle = \begin{pmatrix} e^{-i\phi} \sin \theta/2 \\ -\cos \theta/2 \end{pmatrix} \quad \text{and} \quad |\uparrow\rangle = \begin{pmatrix} e^{-i\phi} \cos \theta/2 \\ \sin \theta/2 \end{pmatrix}$$

These states play the role of our $|n(\lambda)\rangle$ that we had in our general derivation. Note, however, that they are not well defined for all values of \vec{B} . When we have $\theta = \pi$, the angular coordinate ϕ is not well defined. This means that $|\downarrow\rangle$ and $|\uparrow\rangle$ don't have well defined phases. This kind of behaviour is typical of systems with non-trivial Berry phase.

We can easily compute the Berry phase arising from these states (staying away from $\theta = \pi$ to be on the safe side). We have

$$\mathcal{A}_\theta = -i \langle \downarrow | \frac{\partial}{\partial \theta} | \downarrow \rangle = 0 \quad \text{and} \quad \mathcal{A}_\phi = -i \langle \downarrow | \frac{\partial}{\partial \phi} | \downarrow \rangle = -\sin^2 \left(\frac{\theta}{2} \right)$$

The resulting Berry curvature in polar coordinates is

$$\mathcal{F}_{\theta\phi} = \frac{\partial \mathcal{A}_\phi}{\partial \theta} - \frac{\partial \mathcal{A}_\theta}{\partial \phi} = -\frac{1}{2} \sin \theta$$

This is simpler if we translate it back to cartesian coordinates where the rotational symmetry is more manifest. It becomes

$$\mathcal{F}_{ij}(\vec{B}) = -\epsilon_{ijk} \frac{B^k}{2|\vec{B}|^3}$$

But this is interesting. It is a magnetic monopole! Of course, it's not a real magnetic monopole of electromagnetism: those are forbidden by the Maxwell equation. Instead it is, rather confusingly, a magnetic monopole in the space of magnetic fields.

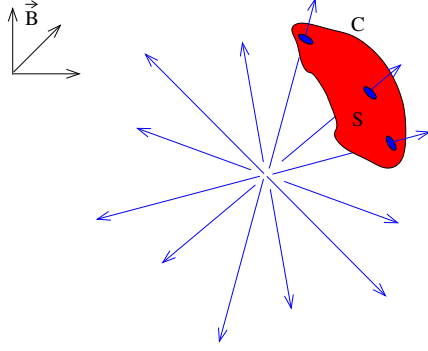


Figure 8: Integrating over S ...

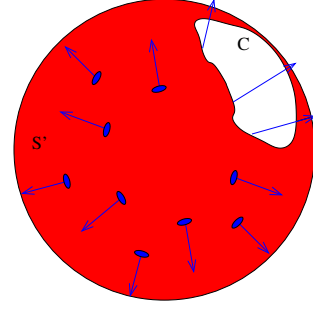


Figure 9: ...or over S' .

Note that the magnetic monopole sits at the point $\vec{B} = 0$ where the two energy levels coincide. Here, the field strength is singular. This is the point where we can no longer trust the Berry phase computation. Nonetheless, it is the presence of this level crossing and the resulting singularity which is dominating the physics of the Berry phase.

The magnetic monopole has charge $g = -1/2$, meaning that the integral of the Berry curvature over any two-sphere \mathbf{S}^2 which surrounds the origin is

$$\int_{\mathbf{S}^2} \mathcal{F}_{ij} dS^{ij} = 4\pi g = -2\pi \quad (1.44)$$

Using this, we can easily compute the Berry phase for any path C that we choose to take in the space of magnetic fields \vec{B} . We only insist that the path C avoids the origin. Suppose that the surface S , bounded by C , makes a solid angle Ω . Then, using the form (1.43) of the Berry phase, we have

$$e^{i\gamma} = \exp \left(-i \int_S \mathcal{F}_{ij} dS^{ij} \right) = \exp \left(\frac{i\Omega}{2} \right) \quad (1.45)$$

Note, however, that there is an ambiguity in this computation. We could choose to form S as shown in the left hand figure. But we could equally well choose the surface S' to go around the back of the sphere, as shown in the right-hand figure. In this case, the solid angle formed by S' is $\Omega' = 4\pi - \Omega$. Computing the Berry phase using S' gives

$$e^{i\gamma'} = \exp \left(-i \int_{S'} \mathcal{F}_{ij} dS^{ij} \right) = \exp \left(\frac{-i(4\pi - \Omega)}{2} \right) = e^{i\gamma} \quad (1.46)$$

where the difference in sign in the second equality comes because the surface now has opposite orientation. So, happily, the two computations agree. Note, however, that this agreement requires that the charge of the monopole in (1.44) is $2g \in \mathbf{Z}$. In the context of electromagnetism, this was Dirac's original argument for the quantisation of

monopole charge. This quantisation extends to a general Berry curvature \mathcal{F}_{ij} with an arbitrary number of parameters: the integral of the curvature over any closed surface must be quantised in units of 2π ,

$$\int \mathcal{F}_{ij} dS^{ij} = 2\pi C \quad (1.47)$$

The integer $C \in \mathbf{Z}$ is called the *Chern number*.

1.5.3 Particles Moving Around a Flux Tube

In our course on *Electromagnetism*, we learned that the gauge potential A_μ is unphysical: the physical quantities that affect the motion of a particle are the electric and magnetic fields. This statement is certainly true classically. Quantum mechanically, it requires some caveats. This is the subject of the Aharonov-Bohm effect. As we will show, aspects of the Aharonov-Bohm effect can be viewed as a special case of the Berry phase.

The starting observation of the Aharonov-Bohm effect is that the gauge potential \vec{A} appears in the Hamiltonian rather than the magnetic field \vec{B} . Of course, the Hamiltonian is invariant under gauge transformations so there's nothing wrong with this. Nonetheless, it does open up the possibility that the physics of a quantum particle can be sensitive to \vec{A} in more subtle ways than a classical particle.

Spectral Flow

To see how the gauge potential \vec{A} can affect the physics, consider the set-up shown in the figure. We have a solenoid of area A , carrying magnetic field \vec{B} and therefore magnetic flux $\Phi = BA$. Outside the solenoid the magnetic field is zero. However, the vector potential is not. This follows from Stokes' theorem which tells us that the line integral outside the solenoid is given by

$$\oint \vec{A} \cdot d\vec{r} = \int \vec{B} \cdot d\vec{S} = \Phi$$

This is simply solved in cylindrical polar coordinates by

$$A_\phi = \frac{\Phi}{2\pi r}$$

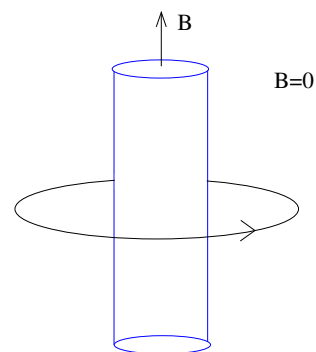


Figure 10: A particle moving around a solenoid.

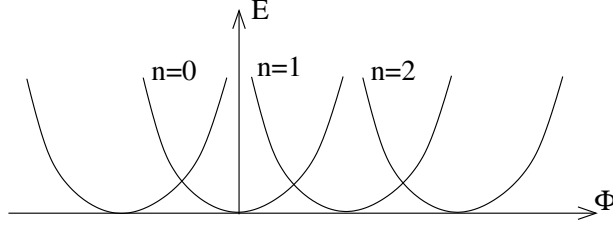


Figure 11: The spectral flow for the energy states of a particle moving around a solenoid.

Now consider a charged quantum particle restricted to lie in a ring of radius r outside the solenoid. The only dynamical degree of freedom is the angular coordinate $\phi \in [0, 2\pi)$. The Hamiltonian is

$$H = \frac{1}{2m} (p_\phi + eA_\phi)^2 = \frac{1}{2mr^2} \left(-i\hbar \frac{\partial}{\partial \phi} + \frac{e\Phi}{2\pi} \right)^2$$

We'd like to see how the presence of this solenoid affects the particle. The energy eigenstates are simply

$$\psi = \frac{1}{\sqrt{2\pi r}} e^{in\phi} \quad n \in \mathbf{Z}$$

where the requirement that ψ is single valued around the circle means that we must take $n \in \mathbf{Z}$. Plugging this into the time independent Schrödinger equation $H\psi = E\psi$, we find the spectrum

$$E = \frac{1}{2mr^2} \left(\hbar n + \frac{e\Phi}{2\pi} \right)^2 = \frac{\hbar^2}{2mr^2} \left(n + \frac{\Phi}{\Phi_0} \right)^2 \quad n \in \mathbf{Z}$$

Note that if Φ is an integer multiple of the quantum of flux $\Phi_0 = 2\pi\hbar/e$, then the spectrum is unaffected by the solenoid. But if the flux in the solenoid is not an integral multiple of Φ_0 — and there is no reason that it should be — then the spectrum gets shifted. We see that the energy of the particle knows about the flux Φ even though the particle never goes near the region with magnetic field. The resulting energy spectrum is shown in Figure 11.

Suppose now that we turn off the solenoid and place the particle in the $n = 0$ ground state. If we increase the flux then, by the time we have reached $\Phi = \Phi_0$, the $n = 0$ state has transformed into the state that we previously labelled $n = 1$. Similarly, each state n is shifted to the next state, $n + 1$. (It is tempting to invoke the adiabatic theorem here but, because of level crossing at $\Phi = \Phi_0/2$ it is not valid.) This is an example of

a phenomenon is called *spectral flow*: under a change of parameter — in this case Φ — the spectrum of the Hamiltonian changes, or “flows”. As we change increase the flux by one unit Φ_0 the spectrum returns to itself, but individual states have morphed into each other. We’ll see related examples of spectral flow applied to the integer quantum Hall effect in Section 2.2.2.

There are actually more lessons lurking in this simple quantum mechanical system. You can read about them in Section 3.6.1 of the [lectures on Gauge Theory](#).

The Aharonov-Bohm Effect

The situation described above smells like the Berry phase story. We can cook up a very similar situation that demonstrates the relationship more clearly. Consider a set-up like the solenoid where the magnetic field is localised to some region of space. We again consider a particle which sits outside this region. However, this time we restrict the particle to lie in a small box. There can be some interesting physics going on inside the box; we’ll capture this by including a potential $V(\vec{x})$ in the Hamiltonian and, in order to trap the particle, we take this potential to be infinite outside the box.

The fact that the box is “small” means that the gauge potential is approximately constant inside the box. If we place the centre of the box at position $\vec{x} = \vec{X}$, then the Hamiltonian of the system is then

$$H = \frac{1}{2m}(-i\hbar\nabla + e\vec{A}(\vec{X}))^2 + V(\vec{x} - \vec{X}) \quad (1.48)$$

We start by placing the centre of the box at position $\vec{x} = \vec{X}_0$ where we’ll take the gauge potential to vanish: $\vec{A}(\vec{X}_0) = 0$. (We can always do a gauge transformation to ensure that \vec{A} vanishes at any point of our choosing). Now the Hamiltonian is of the kind that we solve in our first course on quantum mechanics. We will take the ground state to be

$$\psi(\vec{x} - \vec{X}_0)$$

which is localised around $\vec{x} = \vec{X}_0$ as it should be. Note that we have made a choice of phase in specifying this wavefunction. Now we slowly move the box in some path in space. In doing so, the gauge potential $\vec{A}(\vec{x} = \vec{X})$ experienced by the particle changes. It’s simple to check that the Schrödinger equation for the Hamiltonian (1.48) is solved by the state

$$\psi(\vec{x} - \vec{X}) = \exp\left(-\frac{ie}{\hbar} \int_{\vec{x}=\vec{X}_0}^{\vec{x}=\vec{X}} \vec{A}(\vec{x}) \cdot d\vec{x}\right) \psi(\vec{x} - \vec{X}_0)$$

This works because when the ∇ derivative hits the exponent, it brings down a factor which cancels the $e\vec{A}$ term in the Hamiltonian. We now play our standard Berry game: we take the box in a loop C and bring it back to where we started. The wavefunction comes back to

$$\psi(\vec{x} - \vec{X}_0) \rightarrow e^{i\gamma} \psi(\vec{x} - \vec{X}_0) \quad \text{with} \quad e^{i\gamma} = \exp \left(-\frac{ie}{\hbar} \oint_C \vec{A}(\vec{x}) \cdot d\vec{x} \right) \quad (1.49)$$

Comparing this to our general expression for the Berry phase, we see that in this particular context the Berry connection is actually identified with the electromagnetic potential,

$$\vec{\mathcal{A}}(\vec{X}) = \frac{e}{\hbar} \vec{A}(\vec{x} = \vec{X})$$

The electron has charge $q = -e$ but, in what follows, we'll have need to talk about particles with different charges. In general, if a particle of charge q goes around a region containing flux Φ , it will pick up an Aharonov-Bohm phase

$$e^{iq\Phi/\hbar}$$

This simple fact will play an important role in our discussion of the fractional quantum Hall effect.

There is an experiment which exhibits the Berry phase in the Aharonov-Bohm effect. It is a variant on the famous double slit experiment. As usual, the particle can go through one of two slits. As usual, the wavefunction splits so the particle, in essence, travels through both. Except now, we hide a solenoid carrying magnetic flux Φ behind the wall. The wavefunction of the particle is prohibited from entering the region of the solenoid, so the particle never experiences the magnetic field \vec{B} . Nonetheless, as we have seen, the presence of the solenoid induces a phase difference $e^{i\gamma}$ between particles that take the upper slit and those that take the lower slit. This phase difference manifests itself as a change to the interference pattern seen on the screen. Note that when Φ is an integer multiple of Φ_0 , the interference pattern remains unchanged; it is only sensitive to the fractional part of Φ/Φ_0 .

1.5.4 Non-Abelian Berry Connection

The Berry phase described above assumed that the ground state was unique. We now describe an important generalisation to the situation where the ground state is N -fold degenerate and remains so for all values of the parameter λ .

We should note from the outset that there's something rather special about this situation. If a Hamiltonian has an N -fold degeneracy then a generic perturbation will break this degeneracy. But here we want to change the Hamiltonian without breaking the degeneracy; for this to happen there usually has to be some symmetry protecting the states. We'll see a number of examples of how this can happen in these lectures.

We now play the same game that we saw in the Abelian case. We place the system in one of the N degenerate ground states, vary the parameters in a closed path, and ask: what state does the system return to?

This time the adiabatic theorem tells us only that the system clings to the particular energy eigenspace as the parameters are varied. But, now this eigenspace has N -fold degeneracy and the adiabatic theorem does not restrict how the state moves within this subspace. This means that, by the time we return the parameters to their original values, the state could lie anywhere within this N -dimensional eigenspace. We want to know how it's moved. This is no longer given just by a phase; instead we want to compute a unitary matrix $U \subset U(N)$.

We can compute this by following the same steps that we took for the Abelian Berry phase. To remove the boring, dynamical phase e^{-iEt} , we again assume that the ground state energy is $E = 0$ for all values of λ . The time dependent Schrödinger equation is again

$$i \frac{\partial |\psi\rangle}{\partial t} = H(\lambda(t)) |\psi\rangle = 0 \quad (1.50)$$

This time, for every choice of parameters λ , we introduce an N -dimensional basis of ground states

$$|n^a(\lambda)\rangle \quad a = 1, \dots, N$$

As in the non-degenerate case, there is no canonical way to do this. We could just as happily have picked any other choice of basis for each value of λ . We just pick one. We now think about how this basis evolves through the Schrödinger equation (1.50). We write

$$|\psi_a(t)\rangle = U_{ab}(t) |n_b(\lambda(t))\rangle$$

with U_{ab} the components of a time-dependent unitary matrix $U(t) \subset U(N)$. Plugging this ansatz into (1.50), we have

$$|\dot{\psi}_a\rangle = \dot{U}_{ab} |n_b\rangle + U_{ab} |\dot{n}_b\rangle = 0$$

which, rearranging, now gives

$$U_{ac}^\dagger \dot{U}_{cb} = -\langle n_a | \dot{n}_b \rangle = -\langle n_a | \frac{\partial}{\partial \lambda^i} | n_b \rangle \dot{\lambda}^i \quad (1.51)$$

We again define a connection. This time it is a *non-Abelian Berry connection*,

$$(\mathcal{A}_i)_{ba} = -i \langle n_a | \frac{\partial}{\partial \lambda^i} | n_b \rangle \quad (1.52)$$

We should think of \mathcal{A}_i as an $N \times N$ matrix. It lives in the Lie algebra $u(N)$ and should be thought of as a $U(N)$ gauge connection over the space of parameters.

The gauge connection \mathcal{A}_i is the same kind of object that forms the building block of Yang-Mills theory. Just as in Yang-Mills theory, it suffers from an ambiguity in its definition. Here, the ambiguity arises from the arbitrary choice of basis vectors $|n_a(\lambda)\rangle$ for each value of the parameters λ . We could have quite happily picked a different basis at each point,

$$|n'_a(\lambda)\rangle = \Omega_{ab}(\lambda) |n_b(\lambda)\rangle$$

where $\Omega(\lambda) \in U(N)$ is a unitary rotation of the basis elements. As the notation suggests, there is nothing to stop us picking different rotations for different values of the parameters so Ω can depend on λ . If we compute the Berry connection (1.52) in this new basis, we find

$$\mathcal{A}'_i = \Omega \mathcal{A}_i \Omega^\dagger - i \frac{\partial \Omega}{\partial \lambda^i} \Omega^\dagger \quad (1.53)$$

This is precisely the gauge transformation of a $U(N)$ connection in Yang-Mills theory. Similarly, we can also construct the *curvature* or *field strength* over the parameter space,

$$\mathcal{F}_{ij} = \frac{\partial \mathcal{A}_i}{\partial \lambda^j} - \frac{\partial \mathcal{A}_j}{\partial \lambda^i} - i[\mathcal{A}_i, \mathcal{A}_j]$$

This too lies in the $u(N)$ Lie algebra. In contrast to the Abelian case, the field strength is not gauge invariant. It transforms as

$$\mathcal{F}'_{ij} = \Omega \mathcal{F}_{ij} \Omega^\dagger$$

Gauge invariant combinations of the field strength can be formed by taking the trace over the matrix indices. For example, $\text{tr } \mathcal{F}_{ij}$, which tells us only about the $U(1) \subset U(N)$ part of the Berry connection, or traces of higher powers such as $\text{tr } \mathcal{F}_{ij} \mathcal{F}_{kl}$. However, the most important gauge invariant quantity is the unitary matrix U determined by the differential equation (1.51).

The solution to (1.51) is somewhat more involved than in the Abelian case because of ordering ambiguities of the matrix \mathcal{A}_i in the exponential: the matrix at one point of parameter space, $\mathcal{A}_i(\lambda)$, does not necessarily commute with the matrix at another point $\mathcal{A}_i(\lambda')$. However, this is a problem that we’ve met in other areas of physics⁸. The solution is

$$U = \mathcal{P} \exp \left(-i \oint \mathcal{A}_i d\lambda^i \right)$$

Here $\mathcal{A}_i \in u(N)$ is an $N \times N$ matrix. The notation \mathcal{P} stands for “path ordering”. It means that we Taylor expand the exponential and then order the resulting products so that matrices $\mathcal{A}_i(\lambda)$ which appear later in the path are placed to the right. The result is the unitary matrix $U \in U(N)$ which tells us how the states transform. This unitary matrix is called the *Berry holonomy*.

The non-Abelian Berry holonomy does not play a role in the simplest quantum Hall systems. But it will be important in more subtle quantum Hall states which, for obvious reasons, are usually called *non-Abelian quantum Hall states*. These will be discussed in Section 4⁹.

⁸See, for example, the discussion of Dyson’s formula in Section 3.1 of the *Quantum Field Theory* notes, or the discussion of rotations in Sections 3.1 and 3.7 of the *Classical Dynamics* lecture notes

⁹There are also examples of non-Abelian Berry holonomies unrelated to quantum Hall physics. I have a soft spot for a simple quantum mechanics system whose Berry phase is the BPS ’t Hooft-Polyakov monopole. This was described in J. Sonner and D. Tong, “*Scheme for Building a ’t Hooft-Polyakov Monopole*”, Phys. Rev. Lett 102, 191801 (2009), [arXiv:0809.3783](#).