

Berry phase, the Quantum Hall effect and Topological phases of matter

Andrea Kouta Dagnino

Abstract

We discuss the importance of geometric phase in condensed matter phenomenology, and in particular in the study of the Hall conductance quantization in the Integer Quantum Hall effect (IQHE), whose generalization will lead us to the concept of Chern, and more generally, topological insulators. We review how graphene and topological semimetals exhibit properties of topological insulators despite the absence of a magnetic field. Finally, we present a general overview of topological QFT and Chern-Simons theory in which we formalize our discussion on the IQHE.

Contents

1 Adiabaticity in Quantum mechanics	2	
1.1 Instantaneous eigenstates	2	6.2 The tight-binding approximation for
1.2 The Adiabatic theorem	3	graphene
1.3 A two-level model	4	33
2 Geometric phase	5	6.3 Peierls substitution
2.1 Berry connection and curvature	5	35
2.2 The Berry connection is a connection . .	7	6.4 The Haldane model
3 The Integer Quantum Hall effect	8	36
3.1 The Classical Hall effect	9	6.5 Edge states in the Haldane model
3.2 The Integer Quantum Hall effect	10	38
3.3 Landau levels	10	7 Topological insulators and the \mathbb{Z}_2 invariant 39
3.3.1 Landau gauge	11	7.1 The Kane-Mele model
3.3.2 Symmetric gauge	13	39
3.4 The Hall resistance with filled Landau		7.2 The SSH model and polyacetylene
levels	16	39
3.5 Chiral edge modes	17	8 Spin systems 39
3.6 The importance of disorder	19	8.1 Coherent spin-states
3.7 Spectral flow	21	39
4 The TKNN invariant	22	8.2 Getting the path integral
4.1 Kubo formula	22	40
4.2 Quantization of magnetic fields on a torus	24	8.3 Weiss-Zumino = Pontryagin
4.3 The Kubo formula and Berry connection	24	42
5 Chern insulators	25	8.4 Heisenberg AF chain: θ -terms
5.1 TKNN invariant in Bloch bands	25	43
5.2 Dirac hamiltonians	27	9 Field theory of topological insulators 46
5.3 The Qi-Wu-Zhang model	30	9.1 The parity anomaly
6 Graphene as a Chern insulator	31	46
6.1 The honeycomb lattice	31	9.2 Generalised first Chern number
		48
		9.3 Nielsen-Ninomiya theorem and chiral
		edge states
		48
		10 The Fractional Quantum Hall effect 48
		10.1 Laughlin's wavefunction
		48
		10.2 The Chern-Simons-Landau-Ginzburg
		action
		48
		10.3 Chern-Simons theory
		48
		10.4 Effective-low energy theory
		48
		10.5 Fractional/anyonic excitations
		48
		10.6 Edge excitations
		48

1 Adiabaticity in Quantum mechanics

At the heart of the geometric phase lies the Adiabatic theorem which roughly states that if the parameters in a non-degenerate Hamiltonian don't vary quickly then the system will remain in one of its instantaneous eigenstates if it starts out in one. In other words, as we let the Hamiltonian vary slowly then the instantaneous eigenstates, the eigenstates of the Hamiltonian at an instant in time, will not cross-over and can be tracked smoothly.

1.1 Instantaneous eigenstates

Let us put this discussion into mathematical terms. Consider the time-independent Schroedinger equation:

$$H(t)|\psi(t)\rangle = E(t)|\psi(t)\rangle \quad (1.1.1)$$

This equation does not immediately look time independent as the name suggests, but our goal will be to find its solutions $|\psi(t)\rangle$ at every instant in time t i.e. we are fixing time t and solving the equation. We refer to such solutions as instantaneous eigenstates of the system.

Now let us look at the time-dependent Schroedinger equation

$$i\hbar \frac{\partial}{\partial t} |\Psi(t)\rangle = H(t) |\Psi(t)\rangle \quad (1.1.2)$$

where $|\Psi(t)\rangle$ must not be confused with the instantaneous eigenstates $|\psi(t)\rangle$. From our previous discussions we should expect the system to remain in the instantaneous eigenstate it starts out in, hence we expect $|\Psi(t)\rangle$ to be equal to $|\psi(t)\rangle$ with a phase factor. We thus introduce ansatz

$$|\Psi(t)\rangle = c(t) \exp\left(-\frac{i}{\hbar} \int_0^t E(t') dt'\right) |\psi(t)\rangle \quad (1.1.3)$$

which is analogous to the ansatz $e^{-iEt/\hbar}|E\rangle$ used for a time-independent Hamiltonian. Equation (1.1.3) can be substituted into the TDSE with the LHS giving

$$i\hbar \frac{\partial}{\partial t} |\Psi(t)\rangle = i\hbar \dot{c}(t) \exp\left(-\frac{i}{\hbar} \int_0^t E(t') dt'\right) |\psi(t)\rangle + E(t) |\Psi(t)\rangle \quad (1.1.4)$$

$$+ i\hbar c(t) \exp\left(-\frac{i}{\hbar} \int_0^t E(t') dt'\right) |\dot{\psi}(t)\rangle \quad (1.1.5)$$

and the RHS yielding:

$$H(t) |\Psi(t)\rangle = E(t) |\Psi(t)\rangle \quad (1.1.6)$$

as a consequence of $|\psi(t)\rangle$ being an instantaneous eigenstate. After some simplification we find that:

$$\dot{c}(t) |\psi(t)\rangle + c(t) |\dot{\psi}(t)\rangle = 0 \quad (1.1.7)$$

Dotting to the left with $\langle\psi(t)|$ on both sides we find that:

$$\dot{c}(t) = -c(t) \langle\psi(t)|\dot{\psi}(t)\rangle \quad (1.1.8)$$

and thus the coefficients $c(t)$ are given by:

$$c(t) = \exp\left(-\int_0^t \langle\psi(t')|\dot{\psi}(t')\rangle dt'\right) \quad (1.1.9)$$

This is a bit worrying, since the coefficient looks like it may be exponentially decaying. However, the integrand

$\langle \psi(t) | \dot{\psi}(t) \rangle$ is actually purely imaginary. To prove this consider:

$$\langle \psi(t) | \dot{\psi}(t) \rangle = \int \psi^* \frac{\partial \psi}{\partial t} d^3 \mathbf{r} \quad (1.1.10)$$

$$= \int \left(\frac{\partial |\psi|^2}{\partial t} - \psi \frac{\partial \psi^*}{\partial t} \right) d^3 \mathbf{r} \quad (1.1.11)$$

$$= \frac{\partial}{\partial t} \left(\int |\psi|^2 d^3 \mathbf{r} \right) - \left(\int \psi^* \frac{\partial \psi}{\partial t} d^3 \mathbf{r} \right)^* \quad (1.1.12)$$

The first term clearly vanishes due to normalization, so we find that:

$$\langle \psi(t) | \dot{\psi}(t) \rangle = - \left(\int \psi^* \frac{\partial \psi}{\partial t} d^3 \mathbf{r} \right)^* = - \langle \psi(t) | \dot{\psi}(t) \rangle^* \quad (1.1.13)$$

implying that $\langle \psi(t) | \dot{\psi}(t) \rangle$ is purely imaginary, as required. We can finally write that:

$$|\Psi(t)\rangle = e^{-\frac{i}{\hbar} \int_0^t E(t') dt'} e^{i \int i \langle \psi | \dot{\psi} \rangle dt'} |\psi(t)\rangle \quad (1.1.14)$$

There is one major mistake that we have glossed over. Indeed this result cannot be right because it would imply that a state always remains in its instantaneous eigenstate at all times, now matter what the Hamiltonian looks like. We have made no further assumptions about the system, and got an extremely strong (and dubious) result. The culprit is the dotting to the left with just one state $|\psi(t)\rangle$, which is not enough to specify $c(t)$ as it only gives one component of a vector equation (1.1.7). Wee should instead dot with all states in the Hilbert space (as one does for example in time-independent perturbation theory).

Luckily, we can still salvage this rather nice result by claiming that in the adiabatic limit (1.1.14) is still approximately correct. This is known as the **Adiabatic theorem** which we will now prove.

1.2 The Adiabatic theorem

We must first make some preliminary assumptions on H . We consider a set of instantaneous eigenstates $\{|\psi_n(t)\rangle\}$ which diagonalizes H at time t . Let's impose the initial condition that the system start out in one of these eigenstates $|\Psi(0)\rangle = |\psi_k(0)\rangle$. Let us also assume that the energy levels near k are non-degenerate and can thus be ordered unambiguously:

$$\dots \leq E_{k-1}(t) < E_k(t) < E_{k+1}(t) \leq \dots \quad (1.2.1)$$

The Adiabatic theorem then states that to a good approximation

$$|\Psi(t)\rangle = e^{i\Theta_k(t)} e^{i\gamma_k(t)} |\psi_k(t)\rangle \quad (1.2.2)$$

where the phases are defined as

$$\Theta_k(t) = -\frac{i}{\hbar} \int_0^t E_k(t') dt' \quad (1.2.3)$$

$$\gamma_k(t) = \int_0^t i \langle \psi_k(t') | \dot{\psi}_k(t') \rangle dt' \quad (1.2.4)$$

provided the Hamiltonian varies slowly. How slowly? Let us expand $|\Psi(t)\rangle$ in the instantaneous eigenstate basis:

$$|\Psi(t)\rangle = \sum_n c_n(t) |\psi_n(t)\rangle \quad (1.2.5)$$

The TDSE reads:

$$i\hbar \sum_n (\dot{c}_n(t)|\psi_n(t)\rangle + c_n(t)|\dot{\psi}_n(t)\rangle) = \sum_n c_n(t)E_n(t)|\psi_n(t)\rangle \quad (1.2.6)$$

where $\frac{d}{dt}|\psi(t)\rangle \equiv |\dot{\psi}(t)\rangle$. Dotting to the left with some instantaneous eigenstate $|\psi_m(t)\rangle$, and using the orthonormality of the $\{|\psi_n(t)\rangle\}$ eigenbasis we find that

$$i\hbar\dot{c}_m(t) = c_m(t)E_m(t) - i\hbar \sum_n \langle \psi_m(t) | \dot{\psi}_n(t) \rangle c_n(t) \quad (1.2.7)$$

$$\implies i\hbar\dot{c}_m(t) = \left(E_m(t) - i\hbar \langle \psi_m(t) | \dot{\psi}_m(t) \rangle \right) c_m(t) + \sum_{n \neq m} c_n(t) \langle \psi_m(t) | \dot{\psi}_n(t) \rangle \quad (1.2.8)$$

In particular, for the case of the k th state we start out in

$$i\hbar\dot{c}_k(t) = \left(E_k(t) - i\hbar \langle \psi_k(t) | \dot{\psi}_k(t) \rangle \right) c_k(t) + \sum_{n \neq k} c_n(t) \langle \psi_k(t) | \dot{\psi}_n(t) \rangle \quad (1.2.9)$$

When the last term is negligible then we recover (1.2.2). To justify this cancellation we note that the initial conditions of the system are:

$$c_m(0) = \delta_{km} \quad (1.2.10)$$

so if the state is slowly varying then the last term should indeed be negligible. To verify this, consider:

$$i\hbar\dot{c}_m(0) = i\hbar \sum_{n \neq m} \langle \psi_m | \dot{\psi}_n \rangle c_n(0) = i\hbar \langle \psi_m | \dot{\psi}_k \rangle \neq 0, \quad m \neq k \quad (1.2.11)$$

which is worrying, the other instantaneous eigenstates already start getting occupied at $t = 0$. To see how big $\langle \psi_m | \dot{\psi}_k \rangle$ we look at

$$H(t)|\psi_n(t)\rangle = E_n(t)|\psi_n(t)\rangle \quad (1.2.12)$$

Differentiating with respect to time we get:

$$\dot{H}|\psi_n\rangle + H|\dot{\psi}_n\rangle = \dot{E}_n|\psi_n\rangle + E_n|\dot{\psi}_n\rangle \quad (1.2.13)$$

Therefore:

$$\langle \psi_k | \dot{H} | \psi_n \rangle + E_k \langle \psi_k | \dot{\psi}_n \rangle = E_n \langle \psi_k | \dot{\psi}_n \rangle \implies \langle \psi_k | \dot{\psi}_n \rangle = \frac{\langle \psi_k | \dot{H} | \psi_n \rangle}{E_n - E_k} \quad (1.2.14)$$

For adiabatic changes this term will be small so the mixing between the instantaneous eigenstates will be insignificant. More precisely, we require the change in H over the time-scales $T = \frac{\hbar}{\Delta E}$ associated with the smallest energy gap $\Delta E = E_{k\pm 1} - E_k$ ¹ to be small relative to H :

$$\left| \frac{\dot{H}}{H} \right| \ll \frac{\Delta E}{\hbar} \quad (1.2.15)$$

This is known as the **Adiabatic approximation**.

1.3 A two-level model

We consider a two level system $\{|1\rangle, |2\rangle\}$ modelled by the hamiltonian $H(t) = \frac{\alpha t}{2}(|1\rangle\langle 1| - |2\rangle\langle 2|)$ with instantaneous energy levels $E_1 = \frac{\alpha t}{2}$ and $E_2 = -\frac{\alpha t}{2}$.

¹this is non-zero since we assumed non-degeneracy in the Hamiltonian's spectrum near k

We see that:

$$|\psi_1(t)\rangle = \exp\left(-\frac{i}{\hbar} \int_0^t E_1(t')dt'\right)|1\rangle = e^{-i\alpha t^2/4\hbar}|1\rangle \quad (1.3.1)$$

$$|\psi_2(t)\rangle = \exp\left(-\frac{i}{\hbar} \int_0^t E_2(t')dt'\right)|2\rangle = e^{i\alpha t^2/4\hbar}|2\rangle \quad (1.3.2)$$

are both exact solutions of the TDSE. There is no coupling between the states $|1\rangle$ and $|2\rangle$ despite them crossing at $t = 0$.

We now complicate our model a bit by introducing off-diagonal elements:

$$H = \begin{pmatrix} \alpha t/2 & H_{12} \\ H_{12}^* & -\alpha t/2 \end{pmatrix} \quad (1.3.3)$$

where H_{12} is small. The energy levels are now:

$$E_{\pm} = \pm \sqrt{|H_{12}|^2 + \frac{\alpha^2 t^2}{4}} \quad (1.3.4)$$

We are interested in $t = 0$ where the energy levels $E_{\pm} = \pm H_{12}$ are quite close for small H_{12} . Here the system oscillates between $|1\rangle$ and $|2\rangle$ with Rabi frequency $\omega_{12} = \frac{|H_{12}|}{\hbar}$. We define $\tau_d = \frac{|H_{12}|}{\alpha}$ as the characteristic time scale in which the initial energy levels coinciding with $H_{12} = 0$ get deflected. For an adiabatic approximation we require:

$$\omega_{12}\tau_d \gg 1 \implies \frac{|H_{12}|^2}{\alpha\hbar} \quad (1.3.5)$$

2 Geometric phase

2.1 Berry connection and curvature

Our goal in this section is to gain a better understanding of the $\gamma(t)$ phase.

Consider a Hamiltonian $H(\mathbf{R})$ depending on a set of parameters $\mathbf{R} = (R^1, R^2, \dots, R^N)$ where $\mathbf{R} \in \mathbb{R}^N$ is some vector in the configuration space. Suppose these parameters change with time tracing a path in \mathbb{R}^N given by $\Gamma(t)$ at time t . Finally, let $|\psi_n(\mathbf{R}(t))\rangle$ be the instantaneous eigenstates at time t satisfying the TISE:

$$H(t)|\psi_n(\mathbf{R}(t))\rangle = E_n(t)|\psi_n(\mathbf{R}(t))\rangle \quad (2.1.1)$$

We can use the chain rule

$$i\langle\psi_n(\mathbf{R}(t'))|\frac{d}{dt'}|\psi_n(\mathbf{R}(t'))\rangle = i\langle\psi_n(\mathbf{R}(t'))|\nabla_{\mathbf{R}}|\psi_n(\mathbf{R}(t'))\rangle \cdot \frac{d\mathbf{R}}{dt'} \quad (2.1.2)$$

This leads to the integral in time $\int_0^t dt'$ transforming into a path integral in configuration space $\int_{\Gamma} d\mathbf{R}$. The phase now reads

$$\gamma_n(t) = i \int_{\Gamma(t)} \langle\psi_n(\mathbf{R})|\nabla_{\mathbf{R}}|\psi_n(\mathbf{R})\rangle \cdot d\mathbf{R} \quad (2.1.3)$$

We see that unlike $\Theta(t)$ which acts like a clock keeping track of t , the geometric phase does not really care about time but just the path Γ traversed by the system in the configuration space. In this sense $\gamma_n(t)$ is a purely geometric quantity, it is a **geometric phase**. As (Berry, 1989) puts it, Θ answers the question “how long have you been away” while γ answers the question “where did you go?”.

We may define the **Berry connection** as the integrand of Equation (2.1.3)

$$\mathbf{A}_{(n)}(\mathbf{R}) = i\langle\psi_n(\mathbf{R})|\nabla_{\mathbf{R}}|\psi_n(\mathbf{R})\rangle \quad (2.1.4)$$

Note that the Berry connection is not invariant under gauge transformations, but it does transform in a rather special way. Let us define a new instantaneous eigenstate $|\psi'_n(\mathbf{R})\rangle = e^{-i\alpha(\mathbf{R})}|\psi_n(\mathbf{R})\rangle$. Then the Berry connection transforms as a vector potential:

$$\mathbf{A}'_{(n)}(\mathbf{R}) = i\langle\psi_n(\mathbf{R})|e^{i\alpha(\mathbf{R})}\nabla_{\mathbf{R}}e^{-i\alpha(\mathbf{R})}|\psi_n(\mathbf{R})\rangle \quad (2.1.5)$$

$$= \mathbf{A}_{(n)}(\mathbf{R}) + \nabla_{\mathbf{R}}\alpha \quad (2.1.6)$$

Therefore, the Berry phase pops out because of the ambiguity in the specific phase of our wave-function. The geometric phase now transforms as

$$\gamma'_n(t) = \gamma_n(t) + \int_{\Gamma(t)} (\nabla_{\mathbf{R}}\alpha) \cdot d\mathbf{R} = \gamma_n(t) + \alpha(\mathbf{R}(t)) - \alpha(\mathbf{R}(0)) \quad (2.1.7)$$

We see that in general the geometric phase is not gauge-invariant and thus cannot be observed experimentally. One important exception occurs when the motion completes a closed loop on configuration space, in which case the geometric phase becomes gauge-invariant and thus observable. A well known example of this is the Aharonov-Bohm effect which we will discuss in the next section.

It is important to note that if the instantaneous eigenstates can be chosen to be real then the geometric phase must vanish due to Equation (1.1.13). This does not contradict the fact that the geometric phase is purely imaginary, indeed γ is imaginary, but it happens to also be null. Another case when the geometric phase vanishes is in 1D configuration space (i.e. when only one parameter changes), since a closed loop in 1D just goes in one direction and back, cancelling itself out.

Continuing our analogy between the electromagnetic field and the Berry connection, it makes sense to define the **Berry curvature**:

$$F_{\mu\nu} = \frac{\partial A_\mu}{\partial R^\nu} - \frac{\partial A_\nu}{\partial R^\mu} \quad (2.1.8)$$

which is gauge-invariant as can be easily verified. In $n \geq 3$ dimensions we can use Stoke's theorem to rewrite the Berry phase using the Berry connection. Consider the following differential 1-form in configuration space

$$\mathbf{A} = A_\mu dR^\mu = \langle\psi_n|d|\psi_n\rangle \quad (2.1.9)$$

where d is the exterior derivative. The definition of the Berry curvature allows us to write it as the exterior derivative of A , yielding

$$\mathsf{F} = d\mathbf{A} = F_{\mu\nu} dR^\mu \wedge dR^\nu \implies \gamma(t) = \oint_{\partial S} \mathbf{A} = \int_S \mathsf{F} \quad (2.1.10)$$

In particular, in 3D we find that

$$\gamma(t) = \oint_{\Gamma} \mathbf{A} \cdot d\mathbf{R} = \iint_{S_\Gamma} (\nabla \times \mathbf{A}) \cdot d\mathbf{a} \quad (2.1.11)$$

The Berry curvature can be expressed equivalently as

$$F_{\mu\nu} = i \left(\langle \partial_\mu \psi_n | \partial_\nu \psi_n \rangle - \langle \partial_\nu \psi_n | \partial_\mu \psi_n \rangle \right) \quad (2.1.12)$$

$$= -2\text{Im} \langle \partial_\mu \psi_n | \partial_\nu \psi_n \rangle \quad (2.1.13)$$

The problem with this expression is that the eigenvectors $|\psi_n(\mathbf{R})\rangle$ are gauge-dependent and may carry a phase-

factor that is not smooth in \mathbf{R} . To simplify matters we can differentiate the TISE

$$H|\psi_n(\mathbf{R})\rangle = E_n|\psi_n(\mathbf{R})\rangle \implies \frac{\partial H}{\partial R^\mu}|\psi_n(\mathbf{R})\rangle = \frac{\partial E_n}{\partial R^\mu}|\psi_n(\mathbf{R})\rangle + (E_n - H)|\partial_\mu\psi_n(\mathbf{R})\rangle \quad (2.1.14)$$

so that

$$\langle\psi_m(\mathbf{R})|\partial_\mu H|\psi_n(\mathbf{R})\rangle = \frac{\partial E_n}{\partial R^\mu}\delta_{nm} + (E_n - E_m)\langle\psi_m(\mathbf{R})|\partial_\mu\psi_n(\mathbf{R})\rangle \quad (2.1.15)$$

Consequently we have that for $m \neq n$:

$$\sum_{m \neq n} \langle\psi_m(\mathbf{R})|\partial_\mu H|\psi_n(\mathbf{R})\rangle\langle\psi_n(\mathbf{R})|\partial_\nu H|\psi_m(\mathbf{R})\rangle = \sum_{m \neq n} (E_n - E_m)^2 \langle\psi_m(\mathbf{R})|\partial_\mu\psi_n(\mathbf{R})\rangle\langle\partial_\nu\psi_n(\mathbf{R})|\psi_m(\mathbf{R})\rangle \quad (2.1.16)$$

$$= (E_n - E_m)^2 \langle\partial_\nu\psi_n(\mathbf{R})|\partial_\mu\psi_n(\mathbf{R})\rangle \quad (2.1.17)$$

since the $n = m$ term vanishes. Similarly

$$\sum_{m \neq n} \langle\psi_n(\mathbf{R})|\partial_\mu H|\psi_m(\mathbf{R})\rangle\langle\psi_m(\mathbf{R})|\partial_\nu H|\psi_n(\mathbf{R})\rangle = (E_n - E_m)^2 \langle\partial_\mu\psi_n(\mathbf{R})|\partial_\nu\psi_n(\mathbf{R})\rangle \quad (2.1.18)$$

This leads to the following expression for the Berry curvature

$$F_{\mu\nu} = i \sum_{m \neq n} \frac{\langle\psi_n(\mathbf{R})|\partial_\mu H|\psi_m(\mathbf{R})\rangle\langle\psi_m(\mathbf{R})|\partial_\nu H|\psi_n(\mathbf{R})\rangle - c.c.}{(E_n - E_m)^2} \quad (2.1.19)$$

or alternatively

$$F_{\mu\nu} = -2\text{Im} \left(\sum_{m \neq n} \frac{\langle\psi_n(\mathbf{R})|\partial_\mu H|\psi_m(\mathbf{R})\rangle\langle\psi_m(\mathbf{R})|\partial_\nu H|\psi_n(\mathbf{R})\rangle}{(E_n - E_m)^2} \right) \quad (2.1.20)$$

This is easier to deal with as it does not differentiate the instantaneous eigenstates, but rather just the Hamiltonian.

2.2 The Berry connection is a connection

There is a nice geometric interpretation of the Berry connection which explains its deep link with differential geometry. It is helpful to first look at the Berry phase in the discrete case, so we consider N states $\{|\psi(\mathbf{R}_i)\rangle\}_{i=1,2,\dots,N}$ arranged in a loop as shown beside. In the context of the adiabatic theorem, these would be the evolution of an initial eigenstate $|\psi(\mathbf{R}_1)\rangle$ of an adiabatically evolving Hamiltonian $H(\mathbf{R})$ as \mathbf{R} moves along some loop in configuration space.

The phase acquired by moving along this loop is given by the sum of the phases acquired in each discrete jump from \mathbf{R}_i to \mathbf{R}_{i+1} . Since the latter phase is given by $-\arg(\langle\psi(\mathbf{R}_{i-1})|\psi(\mathbf{R}_i)\rangle)$ we find that

$$\gamma = -\arg(\langle\psi(\mathbf{R}_1)|\psi(\mathbf{R}_2)\rangle\langle\psi(\mathbf{R}_2)|\psi(\mathbf{R}_3)\rangle\dots\langle\psi(\mathbf{R}_{N-1})|\psi(\mathbf{R}_N)\rangle\langle\psi(\mathbf{R}_N)|\psi(\mathbf{R}_1)\rangle) = -\sum_{i=1}^N \arg(\langle\psi(\mathbf{R}_i)|\psi(\mathbf{R}_{i+1})\rangle)$$

where we impose periodic boundary conditions $|\psi(\mathbf{R}_{N+1})\rangle = |\psi(\mathbf{R}_1)\rangle$. It is crucial to note that this quantity is gauge-invariant, for any phase shift by α on any $|\psi(\mathbf{R}_i)\rangle$ would produce $e^{i\alpha}|\psi(\mathbf{R}_i)\rangle$ subsequently cancelled by $e^{-i\alpha}\langle\psi(\mathbf{R}_i)|$. Now suppose that $\arg(\langle\psi'(\mathbf{R}_j)|\psi(\mathbf{R}_{j+1})\rangle) = \gamma_j$ for $2 \leq j < N$. Then by the clever gauge transformation $|\psi(\mathbf{R}_{j+1})\rangle \mapsto |\psi'(\mathbf{R}_{j+1})\rangle \equiv e^{-i\gamma_j}|\psi(\mathbf{R}_{j+1})\rangle$ we can make this phase vanish. The $j = N$ case

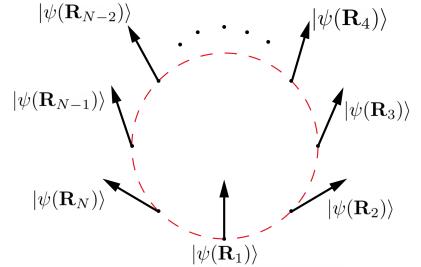


Figure 1: Parallel transport of a state around a path in \mathbf{R} -space.

is an exception: we cannot gauge transform $|\psi(\mathbf{R}_1)\rangle$ as this would get us into an infinite recursive loop of arguments where we would then have to redefine $|\psi(\mathbf{R}_2)\rangle$ etc... Hence all but one phase will be gauged away, leaving a total accumulated phase of

$$\gamma = -\arg(\langle \psi(\mathbf{R}_N) | \psi(\mathbf{R}_1) \rangle) \quad (2.2.1)$$

This gauge fixing procedure is equivalent to a parallel transport condition where there is no phase between adjacent states except at the ends of the loop. The corresponding gauge is known as the **parallel transport gauge**.

Going to the continuum limit we must make the replacement

$$\langle \psi(\mathbf{R}_j) | \psi(\mathbf{R}_{j+1}) \rangle \rightarrow \langle \psi(\mathbf{R}(t)) | \psi(\mathbf{R}(t + \delta t)) \rangle \quad (2.2.2)$$

Assuming our system is adiabatic then we can perform a Taylor expansion to first order in δt

$$|\psi(\mathbf{R}(t + \delta t))\rangle \approx |\psi(\mathbf{R}(t))\rangle + \frac{d|\psi(\mathbf{R}(t))\rangle}{dt} \delta t \quad (2.2.3)$$

implying that

$$\langle \psi(\mathbf{R}(t)) | \psi(\mathbf{R}(t + \delta t)) \rangle = 1 + \langle \psi(\mathbf{R}(t)) | \nabla_{\mathbf{R}} | \psi(\mathbf{R}(t)) \rangle \cdot \delta \mathbf{R} \quad (2.2.4)$$

Also letting $z = \langle \psi(\mathbf{R}(t)) | \nabla_{\mathbf{R}} | \psi(\mathbf{R}(t)) \rangle \cdot \delta \mathbf{R}$ then $\arg(1 + z) = -i \log(1 + z)$ since our states are properly normalized. Consequently using the fact that $z \ll 1$ we may perform a first order Taylor expansion

$$\log(1 + z) \approx z \implies \arg(\langle \psi(\mathbf{R}_j) | \psi(\mathbf{R}_{j+1}) \rangle) = -i \langle \psi(\mathbf{R}(t)) | \nabla_{\mathbf{R}} | \psi(\mathbf{R}(t)) \rangle \cdot \delta \mathbf{R} \quad (2.2.5)$$

We then find that:

$$\gamma(t) = - \sum \arg(\langle \psi(\mathbf{R}(t)) | \nabla_{\mathbf{R}} | \psi(\mathbf{R}(t)) \rangle \cdot \delta \mathbf{R}) \rightarrow i \oint_{\Gamma} \langle \psi(\mathbf{R}(t)) | \nabla_{\mathbf{R}} | \psi(\mathbf{R}(t)) \rangle \cdot d\mathbf{R} \quad (2.2.6)$$

This is the geometric phase we derived earlier! We now see why the Berry phase along a closed loop is invariant, it is the continuum limit of a discrete phase we showed to be gauge invariant. This also means that we cannot simply remove the Berry phase by a clever gauge transformation (just like we could not remove all phases in the discrete case), however we can do so at a particular time (just like we could remove the phase at a particular $n \rightarrow n + 1$ jump). The parallel transport gauge condition can be written as

$$\langle \psi'(\mathbf{R}(t)) | \nabla_{\mathbf{R}} | \psi'(\mathbf{R}(t)) \rangle = 0 \quad (2.2.7)$$

which is satisfied by performing the gauge transformation (2.1.5) with $\alpha = \gamma_n(t)$:

$$|\psi(\mathbf{R}(t))\rangle \mapsto |\psi'(\mathbf{R}(t))\rangle \equiv e^{i\gamma(t)} \psi(\mathbf{R}(t)), \quad \gamma(t) = i \int_{\Gamma(t)} \mathbf{A}(\mathbf{R}) \cdot d\mathbf{R} \quad (2.2.8)$$

Note once again that one could gauge away geometric phase at every point in time except for when the system performs a closed loop in configuration space, in which case any phase change in the wave-function would be cancelled out by an opposite contribution from $e^{i\gamma(t)}$. http://www.physics.iisc.ac.in/~aveek_bid/wp-content/uploads/2019/07/Vanderbilt_ch-3.pdf

3 The Integer Quantum Hall effect

We can now use our pretty framework of geometric phases to study a physical effect, the integer quantum hall effect.

3.1 The Classical Hall effect

The electron transport in a metal was first modelled successfully by Drude in 1900. In this model the atomic lattice gave up one of its valence electrons for each atom, forming a dilute gas of electrons. Ignoring the Coulomb repulsion between electrons, the equation of motion for this gas of electrons in the presence of an electric field \mathbf{E} and magnetic field \mathbf{B} is

$$\left\langle \frac{d\mathbf{v}}{dt} \right\rangle = -\frac{e}{m_e} \mathbf{E} - \frac{e}{m_e} \langle \mathbf{v} \rangle \times \mathbf{B} - \frac{1}{\tau} \langle \mathbf{v} \rangle \quad (3.1.1)$$

where τ is a characteristic viscous damping time which encodes the average collision time between the electrons and the lattice atoms and m_e is the electron mass.

In the steady state $\langle \frac{d\mathbf{v}}{dt} \rangle = 0$ and so we find that the drift velocity is given by:

$$\mathbf{E} = -\langle \mathbf{v} \rangle \times \mathbf{B} - \frac{m_e}{e\tau} \langle \mathbf{v} \rangle \quad (3.1.2)$$

Component by component this reads:

$$E_x = -\frac{m_e}{e\tau} \langle v_x \rangle - \langle v_y \rangle B \quad (3.1.3)$$

$$E_y = \langle v_x \rangle B - \frac{m_e}{e\tau} \langle v_y \rangle \quad (3.1.4)$$

$$E_z = -\frac{m_e}{e\tau} \langle v_z \rangle \quad (3.1.5)$$

where we set our axes so that $\mathbf{B} = B\mathbf{e}_z$. Using the current density $\mathbf{J} = -ne\langle \mathbf{v} \rangle$ we then recover Ohm's law:

$$\mathbf{E} = \overset{\leftrightarrow}{\rho} \mathbf{J} \quad (3.1.6)$$

where we defined the resistivity tensor $\overset{\leftrightarrow}{\rho}$

$$\overset{\leftrightarrow}{\rho} = \begin{pmatrix} \sigma^{-1} & \rho_{xy} & 0 \\ \rho_{yx} & \sigma^{-1} & 0 \\ 0 & 0 & \sigma^{-1} \end{pmatrix}, \quad \sigma = \frac{ne^2\tau}{m_e} \quad (3.1.7)$$

Here $\sigma = \frac{ne^2\tau}{m_e}$ is known as the Drude conductivity, while the off-diagonal component $\rho_{yx} = -\rho_{xy} = -\frac{B}{ne}$ is known as the Hall resistivity. We can invert this relation to find that

$$\mathbf{J} = \overset{\leftrightarrow}{\sigma} \mathbf{E} \quad (3.1.8)$$

where we defined the conductivity matrix

$$\overset{\leftrightarrow}{\sigma} = \frac{1}{1 + \sigma^2 \rho_{xy}^2} \begin{pmatrix} \sigma & -\sigma^2 \rho_{xy} & 0 \\ \sigma^2 \rho_{xy} & \sigma & 0 \\ 0 & 0 & 1 + \sigma^2 \rho_{xy}^2 \end{pmatrix} \quad (3.1.9)$$

Hence if an electric field is applied in the x -direction so as to induce a current J_x in this direction, the off-diagonal components will also produce a transverse electric field $E_y = -\frac{B}{ne\sigma} J_x$. We can define the Hall coefficient as

$$R_H = \frac{\rho_{xy}}{B} \quad (3.1.10)$$

which in the Drude model is $R_H = \frac{1}{ne}$.

From a classical standpoint this makes perfect sense, if we have a flow of electrons moving longitudinally then a perpendicular magnetic field will deflect this flow causing a build up of electrons on one end of the conductor. This produces the transversal electric field, and thus a voltage difference between the two ends of a conductor, known as the Hall voltage. Conversely, if we apply a transversal electric field along x , then this will produce a longitudinal current along y .

3.2 The Integer Quantum Hall effect

At its core, the integer quantum hall effect is a story of particles in. When a strong magnetic field is applied to a metal with freely flowing electrons, the Hall resistivity is measured to be quantized, forming plateaux and occasional spikes as the magnetic field is changed.

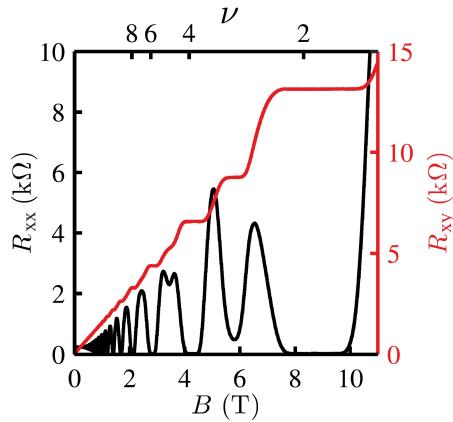


Figure 2: Quantized Hall conductance in a Hall bar with temperature $T = 1.9\text{K}$. Image taken from (Suddards *et al.*, 2012).

On these plateaux the resistivity reads:

$$\rho_{xy} = \frac{2\pi\hbar}{e^2} \frac{1}{\nu}, \quad \nu \in \mathbb{Z} \quad (3.2.1)$$

where ν was measured to be an integer to extraordinary accuracy, up to 10 orders of magnitude! The centers of these plateaux are similarly given by quantized magnetic fields

$$B = \frac{n}{\nu} \Phi_0 \quad (3.2.2)$$

where n is the electron density and $\Phi_0 = \frac{2\pi\hbar}{e}$ is known as the flux quantum which we will soon discuss. A simple calculation can show that this should be the case when exactly ν “Landau levels” (the quantized energy levels in a uniform magnetic field) are filled, but the surprise lies in the fact that the plateaux persist for a range of magnetic fields. Interestingly, the reason for this persistence will turn out to be disorder: the unavoidable impurities in samples used in experiments yield these integers ν .

The longitudinal resistivity ρ_{xx} also presents a surprise, it is vanishing on the plateaux. This would prompt us to define this system as a perfect conductor since $\sigma_{xx} = \frac{1}{\rho_{xx}}$. However, this relation only holds when $\rho_{xy} = 0$, that is in the absence of a magnetic field. When B is turned on we find that $\sigma_{xx} = \frac{\rho_{xx}}{\rho_{xx}^2 + \rho_{xy}^2} \rightarrow 0$. So while there is no longitudinal current flowing ($\sigma_{xx} = 0$) there is also no energy dissipation ($\rho_{xx} = 0$).

3.3 Landau levels

We now turn to the problem of quantizing the motion of particles in a magnetic field. We will ignore any spin degrees of freedom throughout the rest of this essay unless otherwise stated. This is justifiable since in the

presence of a magnetic field there will be a Zeeman splitting between the spin-up and spin-down states of $\Delta = 2\mu_B B$. For large B this splitting is very large so a lot of energy will be required to flip a spin. Assuming we work in the low-energy limit our electrons will be effectively spin-less.

The starting point is a particle of mass m and charge q in a region of space \mathcal{R} with magnetic field \mathbf{B} . In classical mechanics this particle can trace circular trajectories with cyclotron frequency $\omega_c = \frac{qB}{mc}$. The Hamiltonian reads:

$$H = \frac{1}{2m}(\mathbf{p} - q\mathbf{A})^2 \quad (3.3.1)$$

As usual, we can promote \mathbf{p} and \mathbf{A} to quantum operators and solve the resulting Schrödinger equation. This should require choosing a gauge, in particular there are two gauges that reveal different symmetries of the problem. The first is the Landau gauge which imposes translational symmetry making linear momentum a good quantum number, the second is the symmetric gauge which imposes rotational symmetry thus making angular momentum a good quantum number.

3.3.1 Landau gauge

We choose the Landau gauge $\mathbf{A} = xB\mathbf{e}_y$. The Hamiltonian then reads:

$$\hat{H} = \frac{1}{2m} \left[(p_y - qBx)^2 + p_x^2 \right] \quad (3.3.2)$$

where we neglected motion along z which is completely decoupled from the $x - y$ motion. Firstly note that $[H, p_y] = 0$ so we have translational invariance along y . This also means that we should seek Bloch state solutions:

$$\psi(x, y) = \psi(x)e^{ik_y y} \quad (3.3.3)$$

Note that we are only allowed to do this because of our clever gauge choice. Therefore:

$$\hat{H}\psi(x, y) = \frac{1}{2m} \left[(\hbar k_y - qBx)^2 + p_x^2 \right] \psi(x, y) \quad (3.3.4)$$

$$= \left[\frac{p_x^2}{2m} + \frac{1}{2}m \left(\frac{qB}{m} \right)^2 \left(x - \frac{\hbar k_y}{qB} \right)^2 \right] \psi(x, y) \quad (3.3.5)$$

$$= \left[\frac{p_x^2}{2m} + \frac{1}{2}m\omega_c^2(x - x_0)^2 \right] \psi(x, y) \quad (3.3.6)$$

where $x_0 = \frac{\hbar k_y}{qB}$. This is the equation for a simple harmonic oscillator! The characteristic length of an oscillator is $d = \sqrt{\frac{\hbar}{m\omega}}$ which for (3.3.6) reads:

$$d = \sqrt{\frac{\hbar m}{qmB}} = \sqrt{\frac{\hbar}{qB}} \equiv l_B \quad (3.3.7)$$

which we define as the magnetic length. This redefines $x_0 = k_y l_B^2$, which can be interpreted as the value of x about which the solutions will oscillate. Note that the momentum along y determines the localization of the wave-function along x . The Landau eigenstates are thus harmonic oscillator solutions travelling as plane waves along y :

$$\psi_{n,k_y}(x, y) = e^{ik_y y} H_n(x - x_0) e^{-(x-x_0)^2/2l_B^2} \quad (3.3.8)$$

The corresponding energy levels are known as Landau levels:

$$E_{n,k_y} = \hbar\omega_c \left(n + \frac{1}{2} \right), \quad \omega_c = \frac{qB}{m} \quad (3.3.9)$$

Surprisingly the plane wave $e^{ik_y y}$ do not contribute to the total energy, each of these energy levels are totally degenerate in k_y . Consequently there will be a tremendously large degeneracy factor for each $|n\rangle$ state corresponding to how many k_y -modes we can fit in \mathcal{R} .

Indeed suppose we work in the bounded region of space $\mathcal{R} = [0, L_x] \times [0, L_y]$. We take periodic boundary conditions along y so that the possible x -quasimomenta will be quantized as

$$k_y = \frac{2\pi n_y}{L_y} \quad (3.3.10)$$

Since our solutions are exponentially localized at x_0 we also need

$$0 \lesssim x_0 = k_y l_B^2 \lesssim L_x \quad (3.3.11)$$

for our solutions to lie within \mathcal{R} . Consequently $k_y > 0$ so we should only take positive n_y . However, n_x cannot be too positive or else x_0 will get too large, larger than L_x . Thus the allowed n_y values must satisfy:

$$0 < n_y \leq \tilde{n}_y \quad (3.3.12)$$

where \tilde{n}_y is the degeneracy for each $|n\rangle$ landau level in \mathcal{R} and should saturate the upper bound (3.3.11). Thus

$$\max_{n_y}(x_0) = \frac{2\pi \tilde{n}_y}{L_y} l_B^2 = L_x \implies \tilde{n}_y = \frac{L_x L_y}{2\pi \frac{\hbar}{qB}} = \frac{\Phi}{\Phi_0} \quad (3.3.13)$$

where $\Phi = BL_x L_y$ is the total magnetic flux through \mathcal{R} and $\Phi_0 = \frac{2\pi\hbar}{q}$ is the quantum of flux, the flux through a plaquette of area $2\pi l_B^2$. This degeneracy factor tells us how many electrons can occupy a given Landau level, it will prove to be very useful formula later on. Note that as we increase the magnetic field, the degeneracy increases linearly with it.

We can now add a uniform longitudinal electric field $\mathbf{E} = E\mathbf{e}_x$ to our calculations by adding a scalar potential $V = -Ex$ to the electromagnetic field Hamiltonian:

$$\hat{H} = \frac{1}{2m} \left[(p_y - qBx)^2 + p_x^2 \right] - qEx \quad (3.3.14)$$

Using the bloch ansatz

$$\psi(x, y) = \psi(x)e^{ik_y y} \quad (3.3.15)$$

we again find that

$$\hat{H}\psi(x, y) = \frac{1}{2m} \left[(\hbar k_y - qBx)^2 + p_x^2 \right] \psi(x, y) - qEx\psi(x, y) \quad (3.3.16)$$

$$= \left[\frac{p_x^2}{2m} + \frac{1}{2}m \left(\frac{qB}{m} \right)^2 \left(x - \frac{\hbar k_y}{qB} \right)^2 - qEx \right] \psi(x, y) \quad (3.3.17)$$

$$= \left[\frac{p_x^2}{2m} + \frac{1}{2}m\omega_c^2 \left(x - x_0 - \frac{mE}{qB^2} \right)^2 - \frac{E}{B}\hbar k_y - \frac{m}{2}\frac{E^2}{B^2} \right] \psi(x, y) \quad (3.3.18)$$

where in the last step we completed the square. Hence the Landau energy levels are splitted, the mammoth

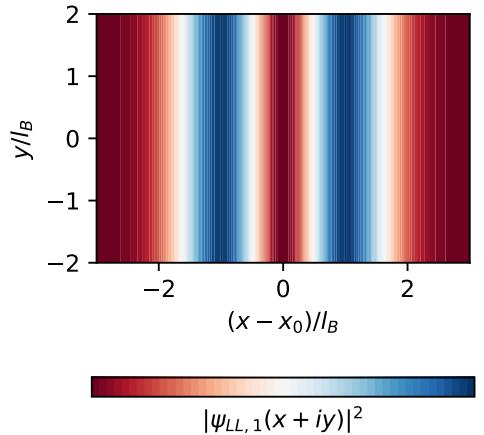


Figure 3: Landau gauge probability amplitude for $n = 0$ (top) and $n = 1$ (bottom)

k -degeneracy has been lifted:

$$E_{n,k_y} = \hbar\omega_c \left(n + \frac{1}{2} \right) - \frac{E}{B} \hbar k_y - \frac{m}{2} \frac{E^2}{B^2} \quad (3.3.19)$$

This time the k_x -modes do contribute to the total energy so we have a “wave-packet” drifting along x . Intuitively this is because the k_x determines where the wave-packet is centered, thus determining the value of the potential energy due to the electric potential. The group velocity of this wave-packet is

$$v_y = \frac{1}{\hbar} \frac{\partial E_{n,k_y}}{\partial k_y} = -\frac{E}{B} \quad (3.3.20)$$

so the particles are drifting along the $\hat{\mathbf{E}} \times \hat{\mathbf{B}} = -\mathbf{e}_y$ direction, just as in the classical prediction!

Setting

$$x'_k = k_y l_B^2 + \frac{qE}{m_e \omega_c^2} = \frac{\hbar k_y}{qB} + \frac{mE}{qB^2} \quad (3.3.21)$$

then we find that the energy levels may be written as

$$E_{n,k_y} = \hbar\omega_c \left(n + \frac{1}{2} \right) - qEx'_k + \frac{1}{2}mv_y^2 \quad (3.3.22)$$

These terms can be clearly interpreted. The first is of course the typical Landau energy. The second term is a potential energy due to the electric potential at the peak of the wave-packet. The final term is a kinetic energy term due to the non-zero drift velocity of the wave-packet.

3.3.2 Symmetric gauge

In the Landau gauge we saw that the harmonic oscillator magically popped out of the Hamiltonian. To do so however we broke translational invariance along y . We now work in the symmetric gauge $\mathbf{A} = \frac{1}{2}B(-y\mathbf{e}_x + x\mathbf{e}_y)$ where we no longer have translational invariance but rather rotational symmetry.

We begin by defining the mechanical momentum

$$\pi = \mathbf{p} - q\mathbf{A} = \left(p_x + \frac{qB}{2}y \right) \mathbf{e}_x + \left(p_y - \frac{qB}{2}x \right) \mathbf{e}_y \quad (3.3.23)$$

and its partner

$$\tilde{\pi} = \mathbf{p} + q\mathbf{A} = \left(p_x - \frac{qB}{2}y \right) \mathbf{e}_x + \left(p_y + \frac{qB}{2}x \right) \mathbf{e}_y \quad (3.3.24)$$

These momenta satisfy the commutation relations:

$$[\pi_x, \pi_y] = -q([p_x, A_y] + [A_x, p_y]) = i\hbar q(\partial_x A_y - \partial_y A_x) = i\hbar qB = i\frac{\hbar^2}{l_B^2} \quad (3.3.25)$$

$$[\tilde{\pi}_x, \tilde{\pi}_y] = q([p_x, A_y] + [A_x, p_y]) = -i\hbar q(\partial_x A_y - \partial_y A_x) = -i\hbar qB = -i\frac{\hbar^2}{l_B^2} \quad (3.3.26)$$

which hold in any gauge, not just the symmetric gauge. Then we can construct the operators

$$a = \frac{1}{\sqrt{2q\hbar B}}(\pi_x + i\pi_y), \quad a^\dagger = \frac{1}{\sqrt{2q\hbar B}}(\pi_x - i\pi_y) \quad (3.3.27)$$

$$b = \frac{1}{\sqrt{2q\hbar B}}(\tilde{\pi}_x - i\tilde{\pi}_y), \quad b^\dagger = \frac{1}{\sqrt{2q\hbar B}}(\tilde{\pi}_x + i\tilde{\pi}_y) \quad (3.3.28)$$

They satisfy the ladder commutation rules

$$[a, a^\dagger] = \frac{1}{2q\hbar B} [\pi_x + i\pi_y, \pi_x - i\pi_y] \quad (3.3.29)$$

$$= \frac{i}{2q\hbar B} ([\pi_y, \pi_x] - [\pi_x, \pi_y]) = 1 \quad (3.3.30)$$

and

$$[b, b^\dagger] = \frac{1}{2q\hbar B} [\tilde{\pi}_x - i\tilde{\pi}_y, \tilde{\pi}_x + i\tilde{\pi}_y] \quad (3.3.31)$$

$$= \frac{i}{2q\hbar B} ([\tilde{\pi}_y, \tilde{\pi}_x] - [\tilde{\pi}_x, \tilde{\pi}_y]) = 1 \quad (3.3.32)$$

Furthermore

$$\hbar\omega_c a^\dagger a = \frac{\hbar q B}{m} \frac{1}{2q\hbar B} ((\mathbf{p} + q\mathbf{A})^2 - i[\pi_x, \pi_y]) = \frac{1}{2m} [(\mathbf{p} + q\mathbf{A})^2 - \hbar q B] = \frac{1}{2m} (\mathbf{p} + q\mathbf{A})^2 - \frac{1}{2} \quad (3.3.33)$$

so we see that the Hamiltonian may be written in terms of a^\dagger, a as a harmonic oscillator

$$H = \hbar\omega_c \left(a^\dagger a + \frac{1}{2} \right) \quad (3.3.34)$$

Note that our arguments so far have been completely gauge-independent.

We can construct the Landau eigenfunctions by converting the ladder operators to differential operators. Restricting ourselves to the Landau gauge we find that

$$a = \frac{1}{\sqrt{2q\hbar B}} \left(-i\hbar \left(\frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right) + \frac{qB}{2}(y - ix) \right) \quad (3.3.35)$$

$$a^\dagger = \frac{1}{\sqrt{2q\hbar B}} \left(-i\hbar \left(\frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right) + \frac{qB}{2}(y + ix) \right) \quad (3.3.36)$$

$$(3.3.37)$$

Letting $z = x + iy$ be a holomorphic coordinate then may define

$$\frac{\partial}{\partial z} = \frac{1}{2} \left(\frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right), \quad \frac{\partial}{\partial \bar{z}} = \frac{1}{2} \left(\frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right) \quad (3.3.38)$$

The ladder operators may then be rewritten as

$$a = \frac{1}{\sqrt{2q\hbar B}} \left(-2i\hbar \frac{\partial}{\partial \bar{z}} - \frac{iqB}{2} z \right) = -\sqrt{2}il_B \left(\frac{\partial}{\partial \bar{z}} + \frac{z}{4l_B^2} \right) \quad (3.3.39)$$

$$a^\dagger = \frac{1}{\sqrt{2q\hbar B}} \left(-2i\hbar \frac{\partial}{\partial z} + \frac{iqB}{2} \bar{z} \right) = -\sqrt{2}il_B \left(\frac{\partial}{\partial z} - \frac{\bar{z}}{4l_B^2} \right) \quad (3.3.40)$$

$$(3.3.41)$$

Similarly we may also write

$$b = \frac{1}{\sqrt{2q\hbar B}} \left(-2i\hbar \frac{\partial}{\partial z} - \frac{iqB}{2} \bar{z} \right) = -\sqrt{2}il_B \left(\frac{\partial}{\partial z} + \frac{\bar{z}}{4l_B^2} \right) \quad (3.3.42)$$

$$b^\dagger = \frac{1}{\sqrt{2q\hbar B}} \left(-2i\hbar \frac{\partial}{\partial z} + \frac{iqB}{2} \bar{z} \right) = -\sqrt{2}il_B \left(\frac{\partial}{\partial \bar{z}} - \frac{z}{4l_B^2} \right) \quad (3.3.43)$$

$$(3.3.44)$$

Now the lowest Landau level states will be annihilated by a so

$$\left(\frac{\partial}{\partial \bar{z}} + \frac{z}{4l_B^2} \right) \psi_{LL} = 0 \implies \psi_{LL} = f(z) e^{-|z|^2/4l_B^2} \quad (3.3.45)$$

where $f(z)$ is some holomorphic functions normalized with Gaussian weight function. Now letting m denote the quantum number associated with b^\dagger/b (we will give it physical significance later) then the lowest Landau state with $m = 0$ will also satisfy

$$\left(\frac{\partial}{\partial z} + \frac{\bar{z}}{4l_B^2} \right) \psi_{LL,m=0} = 0 \quad (3.3.46)$$

Hence

$$\left(\frac{\partial f}{\partial z} - f \frac{\bar{z}}{4l_B^2} + \frac{\bar{z}}{4l_B^2} f \right) e^{-|z|^2/4l_B^2} = 0 \implies f(z) = \text{const.} \quad (3.3.47)$$

so that

$$\psi_{LL,m=0} = A e^{-|z|^2/4l_B^2} \quad (3.3.48)$$

for some normalization constant A . We can increase m by repeatedly acting b^\dagger on this state:

$$\psi_{LL,m} = A_m \left(\frac{z}{l_B} \right)^m e^{-|z|^2/4l_B^2} \quad (3.3.49)$$

Thus the lowest Landau states may be described as skewed Gaussian-like holomorphic functions with a peak at $|z| = \sqrt{2ml_B}$. To see why, note that by the chain rule

$$\frac{\partial}{\partial |z|^2} = \frac{1}{\bar{z}} \frac{\partial}{\partial z} + \frac{1}{z} \frac{\partial}{\partial \bar{z}} = \frac{1}{2|z|} \frac{\partial}{\partial |z|} \quad (3.3.50)$$

so that

$$\frac{\partial}{\partial |z|} \psi = 0 \implies \frac{\partial}{\partial z} \psi = -\frac{\bar{z}}{z} \frac{\partial}{\partial \bar{z}} \psi \quad (3.3.51)$$

Inserting the m th lowest landau level we find that

$$m \frac{z^{m-1}}{l_B^m} - \frac{\bar{z}}{4l_B^2} \left(\frac{z}{l_B} \right)^m = \frac{\bar{z}}{z} \left(\frac{z}{l_B} \right)^m \frac{z}{4l_B^2} e^{-|z|^2/4l_B^2} \implies |z| = \sqrt{2ml_B} \quad (3.3.52)$$

as desired. Intuitively, if m determines the radius of the Landau orbits then it should be the quantum angular momentum number. Indeed, defining the angular momentum operator to be

$$J = i\hbar \left(x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right) = \hbar \left(z \frac{\partial}{\partial z} - \bar{z} \frac{\partial}{\partial \bar{z}} \right) \quad (3.3.53)$$

then it follows immediately that

$$J \psi_{LL,m} = \hbar m \psi_{LL,m} \quad (3.3.54)$$

Finally, we introduce the center of mass position operators

$$X = x + \frac{\pi_y}{m\omega_c}, Y = y - \frac{\pi_x}{m\omega_c} \quad (3.3.55)$$

Then we have that

$$[X, Y] = \frac{1}{m\omega_c} \left(-[x, \pi_x] - [y, \pi_y] + \frac{1}{m\omega_c} [\pi_x, \pi_y] \right) \quad (3.3.56)$$

Note that

$$[x, \pi_x] = [x, p_x - qA_x] = i\hbar, [y, \pi_y] = [y, p_y - qA_y] = i\hbar \quad (3.3.57)$$

so that using $[\pi_x, \pi_y] = i\hbar qB$

$$[X, Y] = -\frac{1}{m\omega_c} (2i\hbar - i\hbar) = -il_B^2 \quad (3.3.58)$$

These coordinates are non-commuting, however they are conserved quantities $[X, H] = [Y, H] = 0$ since

$$[X, \pi_x] = [x + \pi_y/m\omega_c, \pi_x] = i\hbar - \frac{i\hbar qB}{m\omega_c} = 0, [X, \pi_y] = [x, \pi_y] = 0 \quad (3.3.59)$$

and similarly for Y .

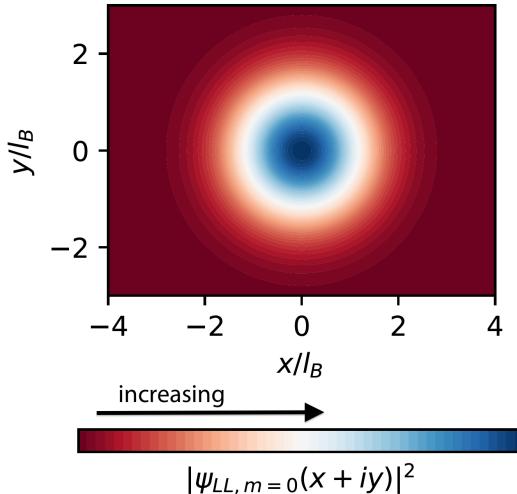


Figure 4: Probability amplitude $|\psi_{LL, m=0}(z)|^2$ of the Lowest landau level with $m = 0$ in the symmetric gauge.

3.4 The Hall resistance with filled Landau levels

Following (Tong, 2016), we now turn to studying how several electrons confined in a Hall system of finite width with a strong magnetic field. For sake of simplicity we will take periodic boundary conditions along y , thus we will consider a Hall cylinder as shown in fig. 5.

Moreover, we can ignore interactions between the electrons and model the many-body states as products of the single-particle states we computed in the previous subsection. We know that classically the “mechanical” momentum for a particle in a magnetic field is given by

$$m\dot{\mathbf{x}} = \mathbf{p} + e\mathbf{A} \quad (3.4.1)$$

where \mathbf{p} is the canonical momentum. Consequently, for an electron moving in a uniform magnetic field the

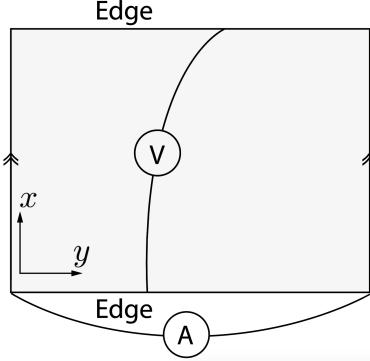


Figure 5: Experimental set-up for the Hall resistance derivation.

current it constitutes is given by $\mathbf{I} = -e\dot{\mathbf{x}}$. In quantum mechanics then we would find that for a collection of electrons

$$\mathbf{I} = -\frac{e}{m} \sum_{\text{occupied}} \langle \psi | -i\hbar\nabla + e\mathbf{A} | \psi \rangle \quad (3.4.2)$$

where $|\psi\rangle$ are the occupied states. We can use the Landau gauge $\mathbf{A} = Bx\mathbf{e}_y$. Then we see that if the first ν Landau levels are filled the transversal current reads

$$I_x = -\frac{e}{m} \sum_{i=1}^{\nu} \sum_{k_y} \langle \psi_{n,k_y} | -i\hbar\partial_x | \psi_{n,k_y} \rangle \quad (3.4.3)$$

However, we know that the expectation value of momentum in a harmonic oscillator state is always zero so $I_x = 0$. On the other hand

$$I_y = -\frac{e}{m} \sum_{i=1}^{\nu} \sum_{k_y} \langle \psi_{n,k_y} | \hbar k_y + eBx | \psi_{n,k_y} \rangle \quad (3.4.4)$$

Again we can use the fact that the harmonic oscillator solutions are localized at $x'_0 = -\frac{\hbar k_y}{eB} - \frac{mE}{eB^2}$ to find that

$$I_y = \nu \sum_{k_y} e \frac{E}{B} = \frac{eAE}{\Phi_0} \nu \quad (3.4.5)$$

where we used the fact that there are $\frac{\Phi}{\Phi_0}$ possible values of k_y localized in the Hall cylinder. The current density is thus given by

$$\mathbf{J} = \begin{pmatrix} 0 \\ eE\nu/\Phi_0 \\ 0 \end{pmatrix} \quad (3.4.6)$$

and using Ohm's law $\mathbf{E} = \overset{\leftrightarrow}{\rho} \mathbf{J}$ we can read off the components of the resistivity tensor

$$\rho_{xx} = \rho_{yy} = \frac{1}{\sigma} = 0 \implies \sigma_{xx} = 0, \rho_{yx} = \frac{E_x}{J_y} = \frac{\Phi_0}{e\nu} = \frac{2\pi\hbar}{e^2} \frac{1}{\nu} \quad (3.4.7)$$

just as we wanted to show!

3.5 Chiral edge modes

One fundamental aspects we have ignored in the derivation of was the behaviour of the electrons at the edge of the Hall cylinder.

Classically, the electrons on the edges will collide with the wall of the Hall cylinder, and since they cannot reverse the direction of their orbit they must skip forward, forming a longitudinal current. Since this occurs at both edges, the net result will be two opposite longitudinal currents that cancel each other out. Electrons confined to move in just one direction are known as chiral, so the two edge states have opposite chiralities.

The quantum explanation is a bit more involved. In the bulk of the system the electron states will be the typical Landau orbits. We assume that the edge of the Hall cylinder can be modelled by a well-like potential $V(x)$ that varies smoothly over length scales $\sim l_B$. Then we can Taylor expand this potential about each state centered at x'_y

$$V(x) \approx V(X) + \frac{\partial V}{\partial x}(x - X) \quad (3.5.1)$$

so ignoring the constant terms we get a linear potential, just as in the electrostatic case! We know how to solve this problem, the drift velocity along the y -direction will be

$$v_y = -\frac{1}{eB} \frac{\partial V}{\partial x} \quad (3.5.2)$$

which is positive on one edge and negative on the other. We therefore have recovered the drifting edge states, this time using a completely quantum mechanical argument!

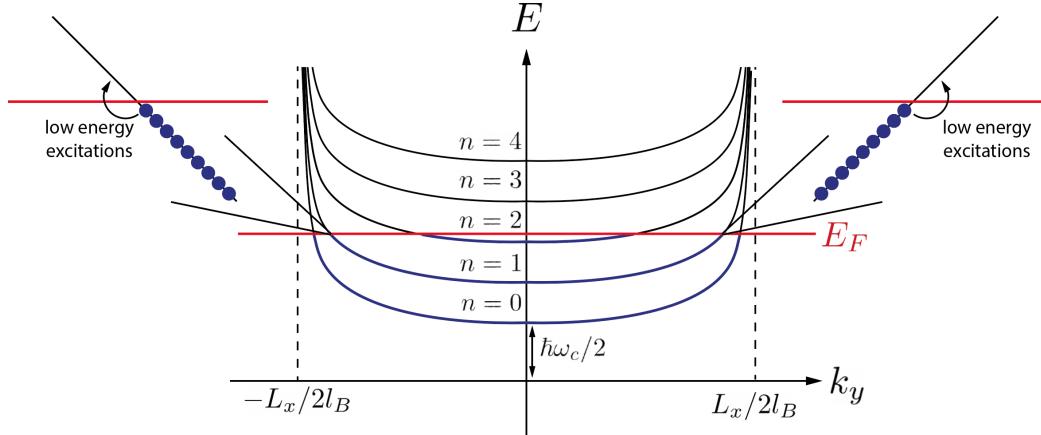


Figure 7: Landau level deformation in a confining potential of the Hall bar. The only possible low-energy excitation occur at the edges of the sample where they constitute edge states.

The situation is shown above.

In order to calculate σ_{xy} we apply a potential difference V_H between the two ends of the Hall cylinder and measure the resulting current I_y . Note that this is equivalent to setting a chemical potential μ_R on the right edge and an elevated chemical potential μ_L on the left edge.

As $T \rightarrow 0$ all states with $E < \mu_R$ are going to be occupied, independent of chirality. The left moving states (left edge) however are also going to occupy states with $\mu_L < E < \mu_R$. Consequently we will have a net

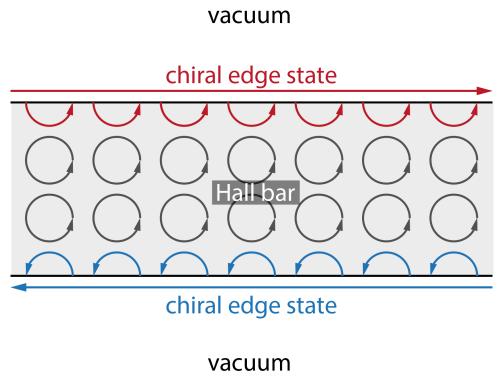


Figure 6: Diagrammatic representation of chiral edge modes with B pointing into the page. The cyclotron orbits in the bulk are unimpeded by the edge of the Hall bar, but the orbits shown in red and blue are “reflected back” forming chiral edge modes.



left-moving current. A $|\psi_{n,k_y}\rangle$ state will take $\frac{L_y}{v_y}$ time to travel along the material so it will constitute a current $\frac{ev_y}{L_y}$. Integrating over all occupied states (we assume for now that only the lowest landau level is occupied):

$$I = -\frac{e}{L_y} \int_{\mu_R}^{\mu_L} dk_y \frac{L_y}{2\pi} v_y = -\frac{e}{2\pi\hbar} \int_{\mu_R}^{\mu_L} dk_y \frac{\partial \epsilon_{k_y}}{\partial k_y} = -\frac{e}{2\pi\hbar} \delta\mu \quad (3.5.3)$$

Defining the Hall voltage drop as:

$$V_H = -\frac{\Delta\mu}{e} \quad (3.5.4)$$

then we have that

$$I = \frac{e^2}{2\pi\hbar} V_H \implies \sigma_{xy} = \frac{e^2}{2\pi\hbar} \quad (3.5.5)$$

This is the required Hall current with $\nu = 1$. Throughout this calculation we have ignored the fact that the Hall voltage will induce a tilt in the Landau levels as shown in fig. 8.

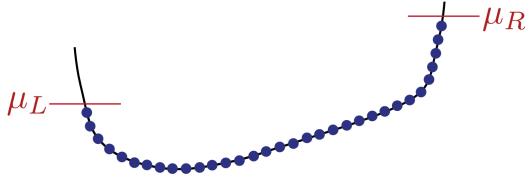


Figure 8: The Landau levels in the presence of a chemical potential gradient.

Luckily our calculation of the current is independent of the shape of the potential as long as it is smooth so that the fundamental theorem of calculus may be applied. We can generalize our argument when ν Landau levels are filled, in which case we will have ν contributions of the type (3.5.5)

$$\sigma_{xy} = \frac{e^2}{2\pi\hbar\nu} \quad (3.5.6)$$

as desired.

3.6 The importance of disorder

The question of why the Hall resistance persists even when the Landau levels are not completely filled has thus far remained a mystery. The answer turns out to lie in the fact that there will always be disorder in whatever experiment we perform, and in the case of the integer quantum Hall effect it is this disorder that makes the resistivity robustly quantized for a range of magnetic fields, with jumps at certain critical magnetic fields.

The impurities in our sample can be modelled by a random potential $V(x, y)$ with $V \ll \hbar\omega_c$ so the troughs and valleys of the disordered potential are smaller than the Landau level gaps. Let us also assume that the potential does not vary on magnetic length scales so $l_B \nabla V \ll \hbar\omega_c$. If we invoke the center of mass coordinate operators

$$X = x + \frac{\pi_y}{m\omega_c}, \quad Y = y - \frac{\pi_x}{m\omega_c} \quad (3.6.1)$$

then it is clear that in the Heisenberg picture they evolve as

$$i\hbar\dot{X} = [X, H + V] = [X, V], \quad i\hbar\dot{Y} = [Y, H + V] = [Y, V] \quad (3.6.2)$$

where we used $[X, H] = [X, Y] = 0$. Since V is smooth over l_B we can use the useful commutator identity:

$$[X, V] = \sum_n V_n [X, Y^n] = \sum_n V_n n[X, Y] Y^{n-1} = [X, Y] \frac{\partial V}{\partial Y} \quad (3.6.3)$$

where we used $[X, Y^n] = n[X, Y] Y^{n-1}$ which can be easily proven by induction. Therefore since $[X, Y] = -il_B^2$ we find the following equations of motion

$$\dot{X} = -\frac{l_B^2}{\hbar} \frac{\partial V}{\partial Y}, \quad \dot{Y} = \frac{l_B^2}{\hbar} \frac{\partial V}{\partial X} \quad (3.6.4)$$

We can interpret these equations as saying that the center of mass will shift along the contour lines of the disordered potential. This makes sense since from our previous discussion of electric fields we know that particles will drift in a direction perpendicular to the applied field. Usually V will have some troughs and valleys, so the particles will orbit these extrema and remain trapped.

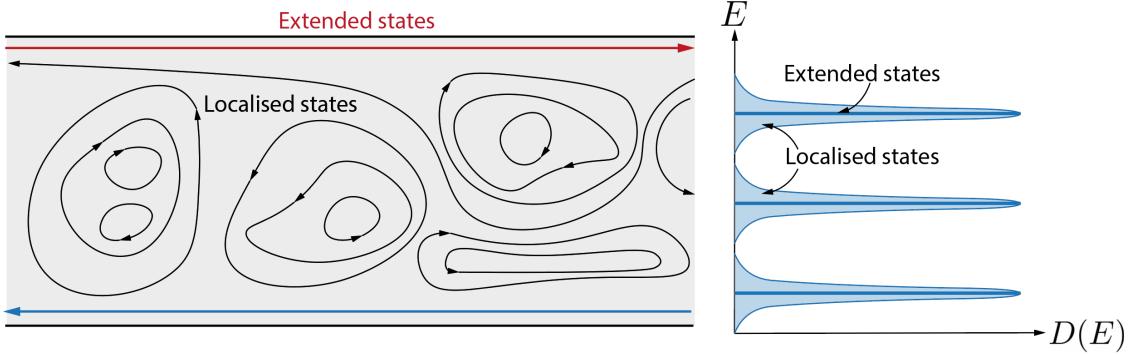


Figure 9: Localised states in the bulk are trapped in either local maxima or minima of the random potential produced by impurities. Thus they will occupy energy levels above or below the degenerate Landau levels. Only the extended states which traverse the entire Hall bar lie at “ground level” and thus occupy the Landau states.

Equipotentials running along the Hall cylinder are thus only possible at the edges where the chiral states are. These will remain localized on their respective edges and not scatter to the other side but are extended over the system. Consequently the density of states will reveal localized states at the extrema of the spectrum (since they lie on troughs or valleys) and extended states at the edges with intermediate energy (since they lie on the “sea level”).

This localization and extension of states is important for conductance since only the latter can contribute to a current. Indeed suppose that we keep the particle number fixed and slowly change the magnetic field. Suppose we start out a full band and start decreasing B . The degeneracy of each Landau level will also decrease, meaning that each band will be able to hold less states. As a result we will stall filling up the localized states in the next band. However, these do not contribute to conduction processes so the conductivity is the same. As long as the magnetic field is not tuned to displace the extended states to localized states or vice-versa (i.e. the chemical potential is not tuned through the extended states) the conductivity will be left unchanged, thus leading to the observed plateaux in σ_{xy} and jumps in σ_{xx} .

We can view this equivalently by a sea-level analogy. As we increase the magnetic field we are populating the localized states, until at some critical magnetic field we finally start occupying the extended states.

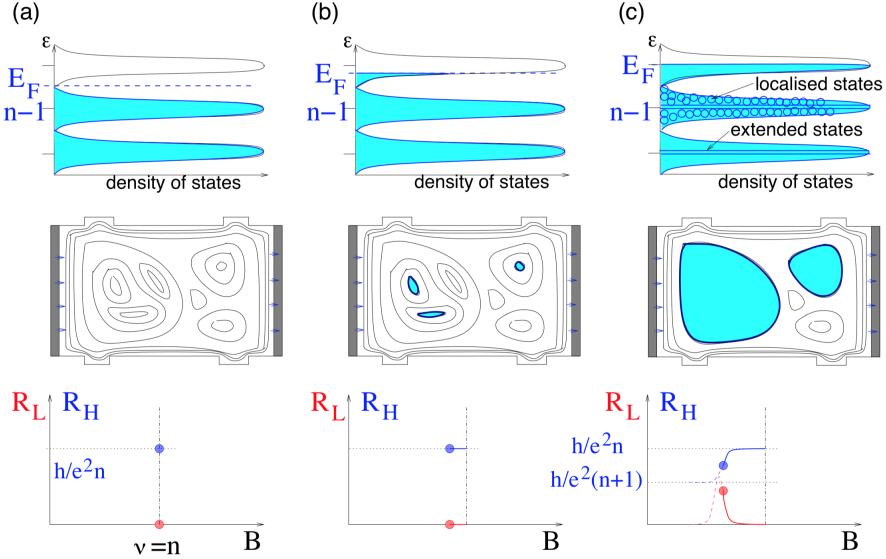


Figure 10: Process of decreasing the magnetic field as seen from a spectral, percolation and hall conductance perspective. In (a) the fermi energy is tuned in a gapped region so all Landau levels are completely filled. As we decrease the magnetic field in (b), each Landau level can hold less states so the localised states start getting occupied. Since these states do not contribute to the conductivity we obtain a plateau. Finally, the localised states now get displaced into extended states and contribute to the Hall conductivity producing a sudden jump. Image taken from (Goerbig, 2009).

Unfortunately this also means that the calculation in the previous two subsections should be wrong since there we assumed that the electrons in the middle of the Hall cylinder also contribute to the current and not just the edge states. Luckily it turns out that the extended states compensate for the localised states by carrying more current.

3.7 Spectral flow

We argue that whatever reason there is behind the Hall quantization it should independent of the shape of the system. Therefore, we may take our Hall bar and deform it into a Corbino ring as shown in fig. 11.

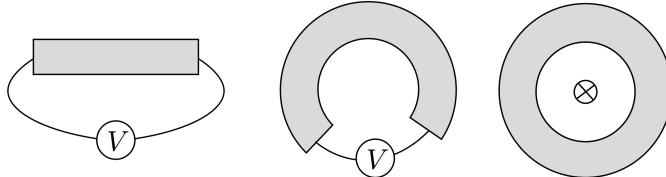


Figure 11: Process of deforming the Hall cylinder in to a Corbino ring.

Now the inner and outer radii form the edges of our sample and host chiral edge states. We can additionally insert a flux tube in the hole whose magnetic flux Φ is easily adjustable. As the flux tube is inserted it will induce an azimuthal electromotive force which represents the applied Hall voltage, and we should expect a radial current to form as a result.

Indeed, suppose we increase Φ from 0 to Φ_0 adiabatically, thus taking a time $\omega_c T \gg 1$. From Lenz's law this

will induce an emf given by

$$\varepsilon = -\frac{d\Phi}{dt} = -\frac{\Phi_0}{T} \quad (3.7.1)$$

Now suppose n electrons have moved from the inner radius to the outer radius forming a radial current $I = -\frac{ne}{T}$, then we have that

$$\rho_{xy} = \frac{\Phi_0}{ne} = \frac{2\pi\hbar}{e^2} \frac{1}{n} \quad (3.7.2)$$

giving the desired quantization. It thus seems reasonable to try to justify our assumption on the electron transport. From our discussion on the Aharonov-Bohm effect we know that as the tube is inserted the Landau states transform as $\psi_m(\Phi = 0) \rightarrow \psi_m(\Phi = \Phi_0) = \psi_{m+1}(\Phi = 0)$ ². Consequently every state will shift from radius $r = \sqrt{2ml_B}$ to $r = \sqrt{2(m+1)l_B}$, the net effect being that an electron is transferred from the inner edge to the outer edge. If n Landau levels are filled then we find that n electrons have effectively moved.

With disorder things get more complicated since we no longer have explicit solutions. Schrödinger's equation reads

$$H = \frac{1}{2m} \left[-\hbar^2 \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial}{\partial r} \right) + \left(-\frac{i\hbar}{r} \frac{\partial}{\partial \phi} + \frac{eBr}{2} + \frac{e\Phi}{2\pi r} \right)^2 \right] + V(r, \theta) \quad (3.7.3)$$

We can try to undo the flux by a gauge transformation

$$\psi(r, \phi) \rightarrow e^{-ie\Phi\phi/2\pi\hbar} \psi(r, \phi) \quad (3.7.4)$$

For the localized states where the wave-function is non-zero only for a small range of ϕ this transformation does not affect the states. However, we must make sure that the chiral edge states, which are extended over the entire ring, are continuous so that $\psi(\phi = 0) \rightarrow \psi(\phi = 2\pi)$. This is true as long as the flux is changed by a multiple of Φ_0 , in which case the Hamiltonian's spectrum will remain unchanged. The only possible change that could have resulted in this is that the states have been shifted and replaced each other, they have undergone **spectral flow**. The explanation is then identical to the case without disorder.

The fact that a system's spectrum is unchanged when a multiple of the quantum flux is added holds more generally in any doubly connected region and for any physical property of the system. This theorem is known as the Byers-Yang theorem.

We have glossed over how exactly this charge pumping can occur. Indeed if the bulk of the ring is localized how does the charge move from one edge to the other? From perturbation theory we know that the localized states are still some superpositions of the Landau states. If we look at the localized states via their superposition then we can see how each Landau state is shifted by spectral flow. In the bulk nothing happens since there is a Landau state from a smaller radius that replaces the Landau state that jumped to the next radius. The only case where this does not happen is at the edges, so the overall effect is that charge is transported radially from one end to the other.

4 The TKNN invariant

We have seen two different explanations for how the Hall conductivity may be quantized, one explanation relies on chiral edge states, while the other lies on a charge pumping argument. The two are related topologically by an invariant discovered in a seminal paper by Thouless, Kohmoto, Nightingale, and den Nijs.

4.1 Kubo formula

We begin by deriving a fundamental result in linear response theory, the Kubo formula.

We consider a hamiltonian H_0 with eigenstates $|m\rangle$ so that $H_0|m\rangle = E_m|m\rangle$. We now add a weak background

²it is tempting to invoke the Adiabatic theorem but this is wrong due to the degeneracy of the Landau levels. The real reason as pointed out earlier is that m is a conserved quantity, so as we increase Φ its value cannot change.

electric field and gauge so that $\mathbf{E} = -\partial_t \mathbf{A}$. This adds an extra perturbation

$$\Delta H = -\mathbf{I} \cdot \mathbf{A} \quad (4.1.1)$$

Although we are ultimately interested in DC electric fields, it will simplify calculations if we take an AC field $\mathbf{E}e^{-i\omega t}$ in the $\omega \rightarrow 0$ limit. Then we have that $\mathbf{A} = \frac{\mathbf{E}}{i\omega} e^{-i\omega t}$, where \mathbf{I} is a current operator.

In the interaction picture an operator O evolves under the base Hamiltonian propagator $O(t) = U_0^{-1} O U_0$ with $U_0 = e^{-iH_0 t/\hbar}$ and the states $|\psi(t_0)\rangle$ evolve under the Dyson propagator $|\psi(t)\rangle = U(t, t_0)|\psi(t_0)\rangle$ where³

$$U(t, t_0) = T \exp \left(-\frac{i}{\hbar} \int_{t_0}^t \Delta H(t') dt' \right) \quad (4.1.2)$$

Since ΔH is a perturbative term we may approximate the propagator to first order as

$$U(t, t_0) = 1 - \frac{i}{\hbar} \int_{t_0}^t \Delta H(t') dt' \quad (4.1.3)$$

Therefore if the system starts in the state $|0\rangle$ at $t = -\infty$ then the expectation value of $\mathbf{I}(t)$ is

$$\langle 0(t) | \mathbf{I}(t) | 0(t) \rangle = \langle 0 | U^\dagger(t) \mathbf{I}(t) U(t) | 0 \rangle \quad (4.1.4)$$

$$= \langle 0 | \left(\mathbf{I}(t) + \frac{i}{\hbar} \int_{-\infty}^t [\Delta H(t'), \mathbf{I}(t)] \right) | 0 \rangle dt' \quad (4.1.5)$$

The first term gives the current in the absence of an electric field and must therefore vanish. Using the fact that $\Delta H(t) = -\frac{I_j E_j}{i\omega} e^{-i\omega t}$ we then find that

$$\langle \mathbf{I}(t) \rangle = \frac{1}{\hbar\omega} \int_{-\infty}^t [I_j(t'), I_k(t)] E_j e^{-i\omega t'} dt' \quad (4.1.6)$$

The correlator can only depend on the difference between t' and t so we can shift our times so that $I_j(t')$ is always evaluated at 0 and thus $I_k(t'')$ is evaluated at times $t'' = t - t'$. This then gives

$$\langle \mathbf{I}(t) \rangle = \frac{1}{\hbar\omega} \int_{-\infty}^t e^{i\omega t''} [I_j(t'), I_k(t)] E_j e^{-i\omega t'} dt'' E_j e^{-i\omega t} \quad (4.1.7)$$

from which we recognise

$$\sigma_{xy} = \frac{1}{L_x L_y} \frac{1}{\hbar\omega} \int_0^\infty e^{i\omega t} \langle 0 | [I_y(0), I_x(t)] | 0 \rangle \quad (4.1.8)$$

Now since $I_x(t) = e^{iH_0 t/\hbar} I_x e^{-iH_0 t/\hbar}$ then we get

$$\sigma_{xy} = \frac{1}{L_x L_y} \frac{1}{\hbar\omega} \int_0^\infty e^{i\omega t} \sum_n (\langle 0 | I_y | n \rangle \langle n | I_x | 0 \rangle e^{i(E_n - E_0)t/\hbar} - \langle 0 | I_x | n \rangle \langle n | I_y | 0 \rangle e^{-i(E_n - E_0)t/\hbar}) \quad (4.1.9)$$

The $n = 0$ term does not contribute so we are left with

$$\sigma_{xy} = -\frac{1}{L_x L_y} \frac{i}{\omega} \sum_{n \neq 0} \left(\frac{\langle 0 | I_y | n \rangle \langle n | I_x | 0 \rangle}{\hbar\omega + E_n - E_0} - \frac{\langle 0 | I_x | n \rangle \langle n | I_y | 0 \rangle}{\hbar\omega - E_n + E_0} \right) \quad (4.1.10)$$

³see my notes on QFT for a proof and discussion of the time ordering operator T

We can expand the fractions in ω and find that

$$\frac{1}{\hbar\omega + E_n - E_0} \approx \frac{1}{E_n - E_0} - \frac{\hbar\omega}{(E_n - E_0)^2} \quad (4.1.11)$$

The first term divided by ω should produce a divergence as $\omega \rightarrow 0$. However, by rotational symmetry we have that $\sigma_{xy} = -\sigma_{yx}$. Then we find that

$$\frac{\langle 0 | I_y | n \rangle \langle n | I_x(0) | 0 \rangle + \langle 0 | I_x | n \rangle \langle n | I_y | 0 \rangle}{E_n - E_0} = -\frac{\langle 0 | I_x | n \rangle \langle n | I_y(0) | 0 \rangle + \langle 0 | I_y | n \rangle \langle n | I_x | 0 \rangle}{E_n - E_0} \quad (4.1.12)$$

The second term on the other hand does not vanish, giving the **Kubo formula** for DC conductance

$$\sigma_{xy} = \frac{i\hbar}{L_x L_y} \sum_{n \neq 0} \frac{\langle 0 | I_y | n \rangle \langle n | I_x | 0 \rangle - \langle 0 | I_x | n \rangle \langle n | I_y | 0 \rangle}{(E_n - E_0)^2} \quad (4.1.13)$$

4.2 Quantization of magnetic fields on a torus

A torus can be formed by imposing periodic boundary conditions along x and y on a rectangle $\mathcal{R} = [0, L_x] \times [0, L_y]$. Let's embed a constant magnetic field \mathbf{B}_0 in the z -direction on this torus.

We must choose a corresponding gauge \mathbf{A} . We see that:

$$B_z = \partial_x A_y - \partial_y A_x = B_0 \implies A_y = B_0 x, A_x = 0 \quad (4.2.1)$$

is a possible gauge, but we must check that it is consistent with the boundary conditions on the torus. Clearly we have that $A_y(x, y + L_y) = A_x(x, y)$, but $A_y(x + L_x, y) \neq A_y(x, y)$. This is fine, since all we need is for $A_y(x + L_x, y)$ to be related by a gauge transformation to $A_y(x, y)$. We need to find this gauge transformation \mathcal{G} :

$$\mathcal{G}A_y(x, y) = A_y(x, y) + \partial_y \Lambda = A_y(x + L_x, y) \implies \partial_y \Lambda = B_0 L_x \implies \Lambda = B_0 L_x y \quad (4.2.2)$$

This gauge parameter seems to be ill-defined at $y = 0 = L_y$, it should take two different values. We can solve this issue by noting that Λ can be ill-defined as long as $e^{iq\Lambda/\hbar}$ is well-defined. We require that:

$$e^{iqB_0L_xy/\hbar c} = e^{iqB_0L_x(y+L_y)/\hbar c} \implies \frac{qB_0L_xL_y}{\hbar} = 2\pi n \quad (4.2.3)$$

so the magnetic field flux should be quantized:

$$\Phi = \frac{2\pi\hbar}{q}n, \quad n \in \mathbb{Z} \quad (4.2.4)$$

The flux quantum is $\Phi_0 = \frac{2\pi\hbar}{q}$ so

$$\Phi = \Phi_0 n, \quad n \in \mathbb{Z} \quad (4.2.5)$$

An alternative derivation consists in defining the magnetic translation operators (gauge dependent, it is easiest to work in the Landau gauge):

$$T(\mathbf{d}) = e^{-i\mathbf{d} \cdot \mathbf{\pi}/\hbar} \quad (4.2.6)$$

and enforcing that moving from $(0, 0)$ to $(L_x, 0)$ and then to (L_x, L_y) is the same as moving from $(0, 0)$ to $(0, L_y)$ and then to (L_x, L_y) :

$$T(L_x \mathbf{e}_x) T(L_y \mathbf{e}_y) = T(L_y \mathbf{e}_y) T(L_x \mathbf{e}_x) \quad (4.2.7)$$

4.3 The Kubo formula and Berry connection

We now consider the same torus as before and add another flux tube Φ_y running azimuthally with it.

The perturbation term now reads:

$$\Delta H = - \sum_{x,y} \frac{I_i \Phi_i}{L_i} \quad (4.3.1)$$

Assuming the states near $|\psi_0\rangle$ are gapped then

$$|\psi_0^{(1)}\rangle = |\psi_0\rangle + \sum_{n \neq \psi_0} \frac{\langle n | \Delta H | \psi_0 \rangle}{E_n - E_0} |n\rangle \quad (4.3.2)$$

so that

$$\frac{d}{d\Phi_i} |\psi_0^{(1)}\rangle = - \frac{1}{L_i} \sum_{n \neq \psi_0} \frac{\langle n | I_i | \psi_0 \rangle}{E_n - E_0} |n\rangle \quad (4.3.3)$$

The sum on the RHS is exactly the term in the Kubo formula:

$$\langle \frac{d\psi_0^{(1)}}{d\Phi_y} | \frac{d\psi_0^{(1)}}{d\Phi_x} \rangle = \frac{1}{L_x L_y} \sum_{n \neq \psi_0} \frac{\langle n | I_x | \psi_0 \rangle \langle \psi_0 | I_y | n \rangle}{(E_n - E_0)^2} \quad (4.3.4)$$

so the Hall conductivity may be written as

$$\sigma_{xy} = i\hbar \left(\langle \frac{d\psi_0^{(1)}}{d\Phi_y} | \frac{d\psi_0^{(1)}}{d\Phi_x} \rangle - \langle \frac{d\psi_0^{(1)}}{d\Phi_x} | \frac{d\psi_0^{(1)}}{d\Phi_y} \rangle \right) \quad (4.3.5)$$

The RHS is of course just the Berry curvature in the flux parameter space! Indeed, letting $\theta_i = \frac{2\pi\Phi_i}{\Phi_0}$ parametrise the flux then the Berry curvature reads

$$F_{xy} = i \left(\langle \frac{d\psi_0^{(1)}}{d\theta_y} | \frac{d\psi_0^{(1)}}{d\theta_x} \rangle - \langle \frac{d\psi_0^{(1)}}{d\theta_x} | \frac{d\psi_0^{(1)}}{d\theta_y} \rangle \right) \quad (4.3.6)$$

Therefore

$$\sigma_{xy} = \frac{e^2}{\hbar} F_{xy} \quad (4.3.7)$$

This of course only takes into account one specific set of values of Φ_x, Φ_y . We can average over the entire torus configuration space \mathbb{T}_Φ of fluxes to find

$$\sigma_{xy} = \frac{e^2}{\hbar} \int_{\mathbb{T}_\Phi} \frac{d^2\theta}{(2\pi)^2} F_{xy} = -\frac{e^2}{2\pi\hbar} C \quad (4.3.8)$$

So the number of Landau levels that are filled which determines the quantization is given by the first Chern number, the IQHE is thus a topological property.

5 Chern insulators

5.1 TKNN invariant in Bloch bands

We now ask ourselves if the topological quantization of the Hall conductivity can also occur in materials without a magnetic field. Such systems are known as Chern insulators, and it turns out that this is indeed possible.

Consider a $L_x \times L_y$ lattice with spacing a along x and spacing b along y . This lattice is in a magnetic field with magnetic periodic boundary conditions, causing the spectrum to fragment into bands. Within each band we have Bloch states in a Brillouin zone labelled by quasimomenta

$$-\frac{\pi}{a} < k_x \leq \frac{\pi}{a}, \quad -\frac{\pi}{b} < k_y \leq \frac{\pi}{b} \quad (5.1.1)$$

and can be expressed as

$$\psi_{n,\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n,\mathbf{k}}(\mathbf{r}) \quad (5.1.2)$$

where for $\mathbf{d}_i = L_y \mathbf{e}_i$

$$T_{\mathbf{d}_i} u_{n,\mathbf{k}}(\mathbf{r}) = e^{i\phi_{\mathbf{d}_i}} u_{n,\mathbf{k}}(\mathbf{r} + \mathbf{d}_i) = u_{n,\mathbf{k}}, \quad (5.1.3)$$

The magnetic Brillouin zone (MBZ) forms a torus \mathbb{T}^2 due to these periodic boundary conditions. As we shall soon see, we cannot cover the entire torus with one chart due to the extra phase factor in the magnetic PBCs. We further assume that the particles are non-interacting and that the spectrum is gapped with the Fermi energy tuned somewhere within a gap, making the system an insulator (according to band theory).

We now claim that each band α has an associated topological invariant known as a **Chern class** C_α arising from the anholonomy of transporting a state along a loop the MBZ. Recall that the Berry connection reads

$$A_i^\alpha(\mathbf{k}) = i \langle u_{n,\mathbf{k}} | \frac{\partial}{\partial k^i} | u_{n,\mathbf{k}} \rangle \quad (5.1.4)$$

giving a Berry curvature

$$F_{ij} = \frac{\partial A_i}{\partial k^j} - \frac{\partial A_j}{\partial k^i} = i \left(\langle \frac{\partial u_{n,\mathbf{k}}}{\partial k_i} | \frac{\partial u_{n,\mathbf{k}}}{\partial k_j} \rangle - \langle \frac{\partial u_{n,\mathbf{k}}}{\partial k_j} | \frac{\partial u_{n,\mathbf{k}}}{\partial k_i} \rangle \right) \quad (5.1.5)$$

The Chern class is then found by integrating F_{xy} over the MBZ

$$C_\alpha = -\frac{1}{2\pi} \int_{\mathbb{T}^2} dk^2 F_{xy} \quad (5.1.6)$$

The importance of the Chern class is that it quantizes the Hall conductance of the lattice

$$\sigma_{xy} = \frac{e^2}{2\pi\hbar} \sum_\alpha C_\alpha \quad (5.1.7)$$

where the sum is taken over all filled bands. This result is known as the **TKNN formula**. We saw how this happened in the Integer quantum hall effect where we had a magnetic field, but we are claiming this occurs in general for non-interacting insulating bands.

To see why C_α should be a topological invariant of our system, let us reconsider the Kubo formula. Since the particles are non-interacting the initial wave-function $|0\rangle$ is going to be a tensor product of single particle states each labelled by α and \mathbf{k} . To find the total conductivity we must therefore sum over the single-particle states in the filled bands. The Kubo formula in the present context then reads

$$\sigma_{xy} = i\hbar \sum_{E_\alpha < E_F < E_\beta} \int_{\mathbb{T}^2} \frac{dk^2}{(2\pi)^2} \frac{\langle u_{\mathbf{k}}^\alpha | I_y | u_{\mathbf{k}}^\beta \rangle \langle n | I_x | 0 \rangle - \langle u_{\mathbf{k}}^\alpha | I_x | u_{\mathbf{k}}^\beta \rangle \langle u_{\mathbf{k}}^\beta | I_y | u_{\mathbf{k}}^\alpha \rangle}{(E_\alpha(\mathbf{k}) - E_\beta(\mathbf{k}))^2} \quad (5.1.8)$$

where α runs over filled bands and β over un-filled bands. The momenta for the two bands should be different but this would lead to cumbersome notation.

We now introduce the modified hamiltonian $\tilde{H}(\mathbf{k}) = e^{-i\mathbf{k}\cdot\mathbf{x}} H e^{i\mathbf{k}\cdot\mathbf{x}}$ so that

$$\tilde{H}(\mathbf{k}) |u_{\mathbf{k}}^\alpha\rangle = E^\alpha(\mathbf{k}) |u_{\mathbf{k}}^\alpha\rangle \quad (5.1.9)$$

For example, if $H = \frac{1}{2m}(\mathbf{p} - q\mathbf{A})^2$ then $\tilde{H} = \frac{1}{2m}(\mathbf{p} - q\mathbf{A} + \hbar\mathbf{k})^2$. Indeed viewing \tilde{H} as H under a change of basis from the right-moving \mathbf{k} -bloch states to the static states u , the momentum operator transforms contravariantly (due to it being a gradient) and thus . Alternatively, since we are moving from a “moving” basis to a “static one” we need to compensate with an extra $\hbar\mathbf{k}$ for the two operators to have the same spectrum.

We thus define the current operator \mathbf{I} as

$$\mathbf{I} = \frac{e}{\hbar} \nabla_{\mathbf{k}} \tilde{H}|_{\mathbf{k}=0} \quad (5.1.10)$$

For the single particle case this reduces to:

$$\mathbf{I} = \frac{e}{2m\hbar} \nabla_{\mathbf{k}} (\mathbf{p} - q\mathbf{A} + \hbar\mathbf{k})^2 = \frac{e}{m\hbar} \hbar\pi \quad (5.1.11)$$

as expected. Now the Kubo formula reads

$$\sigma_{xy} = \frac{ie^2}{\hbar} \sum_{\alpha, \beta} \int_{\mathbb{T}^2} \frac{dk^2}{(2\pi)^2} \frac{\langle u_{\mathbf{k}}^{\alpha} | \partial_y \tilde{H} | u_{\mathbf{k}}^{\beta} \rangle \langle n | \partial_x \tilde{H} | 0 \rangle - \langle u_{\mathbf{k}}^{\alpha} | \partial_x \tilde{H} | u_{\mathbf{k}}^{\beta} \rangle \langle u_{\mathbf{k}}^{\beta} | \partial_y \tilde{H} | u_{\mathbf{k}}^{\alpha} \rangle}{(E_{\alpha}(\mathbf{k}) - E_{\beta}(\mathbf{k}))^2} \quad (5.1.12)$$

but we know from our discussion of the Berry curvature that (2.1.19)

$$F_{\mu\nu} = i \sum_{\beta \neq \alpha} \frac{\langle u_{\mathbf{k}}^{\alpha} | \partial_{\mu} \tilde{H} | u_{\mathbf{k}}^{\beta} \rangle \langle u_{\mathbf{k}}^{\alpha} | \partial_{\nu} \tilde{H} | u_{\mathbf{k}}^{\beta} \rangle - c.c.}{(E_{\alpha}(\mathbf{k}) - E_{\beta}(\mathbf{k}))^2} \quad (5.1.13)$$

where again the momenta for α and β should be different. Consequently we find that

$$\sigma_{xy} = -\frac{e^2}{2\pi\hbar} \sum_{\alpha} \int_{\mathbb{T}^2} dk^2 F_{xy} = \frac{e^2}{2\pi\hbar} \sum_{\alpha} C_{\alpha} \quad (5.1.14)$$

which is exactly what we wanted to show! When an integer number of bands are filled the Hall conductance of a Chern insulator is quantized by its (first) Chern number. Importantly, this formula does not depend on the geometry of the system, if we deform it continuously the conductance will remain the same due to topological protection.

5.2 Dirac hamiltonians

We try to construct the simplest possible Chern insulator. We of course need two bands, one below and one above the band gap in which we tune in the Fermi energy. A two-state system will do the job, so we consider a two-dimensional Hilbert space $\mathcal{H} \cong \mathbb{C}^2$ and a Hamiltonian $H \in U(2)$ of the form

$$\tilde{H} = \begin{pmatrix} h_0 + h_z & h_x - ih_y \\ h_x + ih_y & h_0 - h_z \end{pmatrix} \quad (5.2.1)$$

which can be written using Pauli matrices as a Dirac Hamiltonian

$$\tilde{H}(\mathbf{k}) = h_{\mu}(\mathbf{k}) \sigma^{\mu} = h(\mathbf{k}) \cdot \boldsymbol{\sigma} + h_0(\mathbf{k}) \mathbb{1} \quad (5.2.2)$$

Note that $\text{tr} \tilde{H} = 2h_0$ and $\det \tilde{H} = h_0^2 + |\mathbf{h}|^2$ implying that the energy levels of the system are

$$E_{\pm}(\mathbf{k}) = h_0(\mathbf{k}) \pm |\mathbf{h}(\mathbf{k})| \quad (5.2.3)$$

with eigenstates

$$|\pm\rangle = \sqrt{\frac{h_x^2 + h_y^2}{E_{\pm}^2 + |\mathbf{h}|^2}} \begin{pmatrix} \frac{E_{\pm}}{h_x + ih_y} \\ 1 \end{pmatrix} \quad (5.2.4)$$

Since h_0 shifts both energies it has no topological importance and can be safely neglected. We can then write in spherical coordinates

$$\mathbf{h}(\theta, \phi) = |\mathbf{h}| \begin{pmatrix} \sin \theta \cos \phi \\ \sin \theta \sin \phi \\ \cos \theta \end{pmatrix}, \quad 0 < \phi \leq 2\pi, \quad 0 \leq \theta \leq \pi \quad (5.2.5)$$

so the configuration space is a sphere \mathbb{S}^2 . The ground state, which is occupied, then reads (ignore the N for the moment)

$$|-\rangle^N(\theta, \phi) = \begin{pmatrix} -\sin(\theta/2) \\ e^{i\phi} \cos(\theta/2) \end{pmatrix} \quad (5.2.6)$$

Unfortunately this eigenvector is not well-defined on the entire sphere, it has a singularity at $\theta = 0$, that is at the North pole, due to an ill-defined phase ϕ ⁴. Indeed, it could have equally been written as (ignore the S for the moment):

$$|-\rangle^S(\theta, \phi) = \begin{pmatrix} -e^{i\phi} \sin(\theta/2) \\ \cos(\theta/2) \end{pmatrix} \quad (5.2.7)$$

which is instead ill-defined on the South-pole where $\theta = \pi$.

We can therefore solve this issue by considering two separate charts (x_N, U_N) and (x_S, U_S) on the north and south hemispheres respectively. Within each chart the states $|-\rangle^S(\theta, \phi)$ and $|-\rangle^N(\theta, \phi)$ are well-defined. The two charts can be glued by imposing the boundary condition

$$|-\rangle^S(\theta, \phi) = e^{i\phi} |-\rangle^N(\theta, \phi) \quad (5.2.8)$$

along the equator $\theta = \pi$. We will now prove that the Berry curvature reads

$$\mathsf{F} = \frac{1}{4} \epsilon_{ijk} \frac{h^i}{|\mathbf{h}|^3} dh^j \wedge dh^k \quad (5.2.9)$$

where d is the exterior derivative which transforms as

$$dh^j = \frac{\partial h^j}{\partial k^n} dk^n \implies dh^j \wedge dh^k = \frac{\partial h^j}{\partial k^n} \frac{\partial h^k}{\partial k^m} dk^n \wedge dk^m \quad (5.2.10)$$

so that

$$C = \frac{1}{2\pi} \int_{\mathbb{T}^2} \mathsf{F} = \frac{1}{2\pi} \int_{\mathbb{T}^2} \frac{1}{4} \epsilon_{ijk} \frac{h^i}{|\mathbf{h}|^3} \frac{\partial h^j}{\partial k^n} \frac{\partial h^k}{\partial k^m} dk^n \wedge dk^m \quad (5.2.11)$$

Recalling that $\mathbf{k} = (k_x, k_y)$ and using the symmetry of the integrand this can be written as

$$C = \frac{1}{4\pi} \int_{\mathbb{T}^2} \frac{\mathbf{h}}{|\mathbf{h}|^3} \cdot \left(\frac{\partial \mathbf{h}}{\partial k^x} \times \frac{\partial \mathbf{h}}{\partial k^y} \right) dk^x \wedge dk^y \quad (5.2.12)$$

There is a nice geometrical interpretation of this formula, it calculates how many times the Brillouin torus \mathbb{T}^2 wraps around the origin, it is thus also known as the **winding number** of the BZ. Because \mathbf{d} is (generally) a smooth mapping, it will map the BZ, which is a compact manifold, onto another compact manifold. This implies that the winding number must be an integer!

This interpretation of the Chern number is visually simple to understand but there are faster ways to actually calculate the C. To actually perform this calculation, we go to the spherical coordinate charts on the two hemispheres. Then we see that

$$A_i^N = i \langle -^N | \partial_{k_i} | -^N \rangle = i \left(\begin{pmatrix} -\sin \theta/2 \\ e^{-i\phi} \cos \theta/2 \end{pmatrix}^T \frac{\partial}{\partial k^i} \begin{pmatrix} -\sin \theta/2 \\ e^{i\phi} \cos \theta/2 \end{pmatrix} \right) \quad (5.2.13)$$

$$= -\cos^2 \frac{\theta}{2} \frac{\partial \phi}{\partial k^i} \quad (5.2.14)$$

⁴this does not occur at $\theta = \pi$ since there the cosine term vanishes

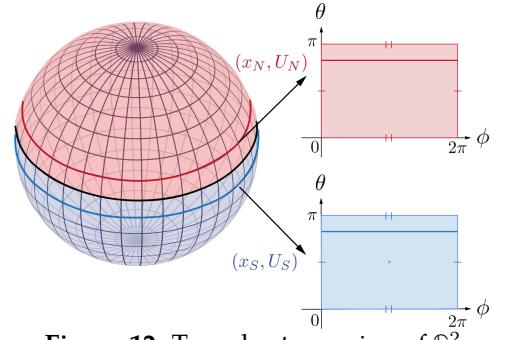
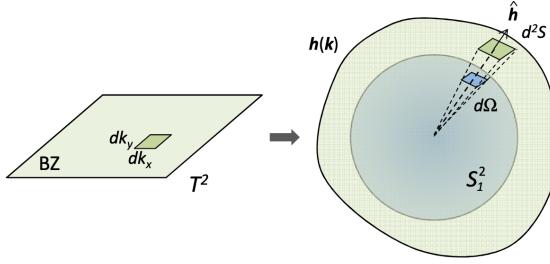


Figure 12: Two-chart covering of \mathbb{S}^2 .



Similarly we find that

$$A_i^S = i \langle -^S | \partial_{k_i} | -^S \rangle = i \left(\frac{-e^{i\phi} \sin \theta / 2}{\cos \theta / 2} \right)^T \frac{\partial}{\partial k^i} \left(\frac{-e^{-i\phi} \sin \theta / 2}{\cos \theta / 2} \right) \quad (5.2.15)$$

$$= \sin^2 \frac{\theta}{2} \frac{\partial \phi}{\partial k^i} \quad (5.2.16)$$

so in differential form notation we have that

$$\mathbf{A}^N = -\cos^2 \frac{\theta}{2} d\phi, \quad \mathbf{A}^S = \sin^2 \frac{\theta}{2} d\phi \quad (5.2.17)$$

The Berry curvature is then the exterior derivative of this

$$\mathbf{F}^N = d\mathbf{A}^N = \frac{\partial}{\partial k^i} (\cos^2 \theta / 2) dk^i \wedge d\phi \quad (5.2.18)$$

$$= \frac{1}{2} \sin \theta \frac{\partial \theta}{\partial k_i} dk^i \wedge d\phi = \frac{1}{2} \sin \theta d\theta \wedge d\phi \quad (5.2.19)$$

Of course, due to the gauge invariance of the Berry curvature we will get the same in the southern hemisphere so we can write

$$\mathbf{F} = \frac{1}{2} \sin \theta d\theta \wedge d\phi \quad (5.2.20)$$

It is clear then that the associated Berry phase along some closed loop Γ is given by

$$\gamma = \exp \left(i \oint_{\Gamma} \mathbf{F} \right) = \exp \left(i \oint_{\Gamma} \frac{1}{2} \sin \theta d\theta d\phi \right) = e^{i\Omega/2} \quad (5.2.21)$$

where Ω is the solid angle subtended by Γ . Similarly

$$C = \frac{1}{4\pi} \int_{\mathbb{T}^2} \sin \theta d\theta \wedge d\phi \quad (5.2.22)$$

can be seen as the ratio between the solid angle subtended by the entire torus and the solid angle of \mathbb{S}^2 , the Chern number counts how many times \mathbb{T}^2 wraps around \mathbb{S}^2 .

We can write \mathbf{F} in Cartesian coordinates by noting that

$$d\theta \wedge d\phi = \frac{\partial \theta}{\partial h^j} \frac{\partial \phi}{\partial h^k} dh^j \wedge dh^k, \quad \text{and} \quad \sin \theta \frac{\partial \theta}{\partial h^j} \frac{\partial \phi}{\partial h^k} = \epsilon_{ijk} h^i \quad (5.2.23)$$

giving

$$\mathbf{F} = \frac{1}{4} \epsilon_{ijk} \frac{h^i}{|\mathbf{h}|^3} dh^j \wedge dh^k \quad (5.2.24)$$

This is the field due to a monopole, a magnetic monopole not in real space but in the configuration space of our Hamiltonian! This gives us one final interpretation of the Chern number for two-level systems: the Berry curvature is a magnetic monopole field centered at the origin in phase space, and the Chern number is thus the flux of this monopole field. This must of course be an integer number if the Brillouin zone is a compact manifold.

5.3 The Qi-Wu-Zhang model

Consider a continuum model with a Dirac hamiltonian where $h_z = m$ and $h_x = k_x, h_y = k_y$ so that

$$\tilde{H} = k_x \sigma_x + k_y \sigma_y + m \sigma_z \quad (5.3.1)$$

It is clear that the Berry curvature now reads

$$F_{xy} = \frac{1}{2} \frac{m}{(k^2 + m^2)^{3/2}} \quad (5.3.2)$$

The Hall conductance is then given by

$$\sigma_{xy} = \frac{e}{\hbar} \frac{1}{2\pi} \int \frac{1}{2} \frac{m}{(k^2 + m^2)^{3/2}} d^2\mathbf{k} = \frac{e}{\hbar} \frac{m}{2} \int_0^\infty \frac{k}{(k^2 + m^2)^{3/2}} dk = \frac{e \operatorname{sgn}(m)}{\hbar} \frac{1}{2} \quad (5.3.3)$$

This result may seem contradictory to the TKNN theorem, but recall that the latter applies to lattice problems only, while here we did not regularize the momenta of the electrons. We therefore have no reason to expect the same integer quantization to occur.

We can consider the lattice model equivalent of the above by defining

$$h_x = \sin k_x, h_y = \sin k_y, h_z = m + 2 - \cos k_x - \cos k_y \quad (5.3.4)$$

so that

$$\tilde{H} = \sin k_x \sigma_x + \sin k_y \sigma_y + (m + 2 - \cos k_x - \cos k_y) \sigma_z \quad (5.3.5)$$

Note that as in the $k \rightarrow 0$ limit this exactly reduces to the continuum model we looked at. The energy levels are

$$E_\pm(m, \mathbf{k}) = \pm \sqrt{\sin^2 k_x + \sin^2 k_y + (m + 2 - \cos k_x - \cos k_y)^2} \quad (5.3.6)$$

so this model is gapped except for four instances when $E_+ = E_- = 0$:

- (i) $m = 0, (k_x, k_y) = (0, 0)$
- (i) $m = -2, (k_x, k_y) = (\pi, 0)$ and $(k_x, k_y) = (0, \pi)$
- (i) $m = -4, (k_x, k_y) = (\pi, \pi)$

For parameters where the gap does not close we will get different topological phases which we can characterize via their Chern number. Indeed when the gap closes our we cannot tune the chemical potential to lie within it, so our discussion of the TKNN invariant breaks down. We should therefore expect topological properties of our system to vary at these points in configuration space. To calculate the Chern numbers we consider starting in the $m \rightarrow \infty$ limit and decreasing m slowly.

If $m > 0$ then we are in a phase topologically equivalent to $m \rightarrow \infty$, that is in the atomic limit. Since the Dirac fermions are localized there will be no Hall conductance, so the Chern number will be zero.

As we continue to decrease m eventually we encounter the gap closure at $m = 0$, and subsequently enter the $-2 < m < 0$ phase. Near $m = 0$ we know that the model looks roughly like the continuum model

$$\tilde{H} \approx k_x \sigma_x + k_y \sigma_y + m \sigma_z \quad (5.3.7)$$

so the total change in the Chern number will be

$$\Delta C = \frac{1}{2}(\operatorname{sgn}(m < 0) - \operatorname{sgn}(m > 0)) = \frac{1}{2}((-1) - 1) = -1 \quad (5.3.8)$$

Since we started with a Chern number of 0 the Chern number in the $-2 < m < 0$ phase will be -1 .

As we continue to decrease m we eventually reach the gap closure at $m = -2$ and subsequently enter the $-4 < m < -2$ phase. The gap closes at two points $\mathbf{k} = (\pi, 0)$ and $\mathbf{k} = (0, \pi)$ so expanding the lattice model near this point

$$\tilde{H}_{(\pi,0)} \approx -k_x\sigma_x + k_y\sigma_y + (m+2)\sigma_z \quad (5.3.9)$$

$$\tilde{H}_{(0,\pi)} \approx k_x\sigma_x - k_y\sigma_y + (m+2)\sigma_z \quad (5.3.10)$$

Note that we have an overall sign change in either k_x or k_y so when performing the integral over the Brillouin zone we will gain an extra minus sign (the azimuthal momentum coordinate $\phi = \arctan(k_x/k_y)$ changes sign, so when integrated we get -2π rather than 2π). The total change in Chern number is then

$$\Delta C = -2 \cdot \frac{1}{2}(\operatorname{sgn}(m+2 < 0) - \operatorname{sgn}(m+2 > 0)) = +2 \quad (5.3.11)$$

Thus the Chern number in this phase is 1.

Finally, we continue decreasing m until we eventually reach the gap closure at $m = -4$ and enter the $m < -4$ phase. We can use the same argument as with the $m > 0$ phase, namely that we are in the atomic limit where the fermions are localized and therefore don't conduct. As a sanity check, we can repeat the change in chern number argument. Expanding the Hamiltonian near $\mathbf{k} = (\pi, \pi)$ we get that

$$\tilde{H} \approx -k_x\sigma_x - k_y\sigma_y + (m+4)\sigma_z \quad (5.3.12)$$

so that

$$\Delta C = \frac{1}{2}(\operatorname{sgn}(m+4 < 0) - \operatorname{sgn}(m+4 > 0)) = -1 \quad (5.3.13)$$

Hence we get a zero chern number as desired.

Alternatively, we can calculate by interpreting the Chern number in (5.2.12) as the winding number of Brillouin zone mapped under $d(\mathbf{k})$. We plot three different surfaces corresponding to the three distinct phases we identified:

6 Graphene as a Chern insulator

We have seen how bands in time-reversal symmetry breaking models can attain non-zero chern number changes during gap closings. Such materials are known as Chern insulators. In non-trivial topological phases, which cannot be adiabatically continued to the atomic limit, chiral edge states emerge at the boundaries of the chern insulator much like in the IQHE. This leads to the quantisation of the Hall conductance even in the absence of a net magnetic flux, and is a phenomenon known as the **anomalous quantum Hall effect**. The goal of this section will be to construct the **Haldane model** (Haldane, 1988), another Chern insulator which arises directly from the tight-binding model of a honeycomb lattice. This will also set the groundwork for further investigations into the **quantum spin Hall effect** first proposed by C.L. Kane and E. Mele (Kane and Mele, 2005).

6.1 The honeycomb lattice

The honeycomb lattice has a special geometry which allows for some interesting physics to occur. Let the

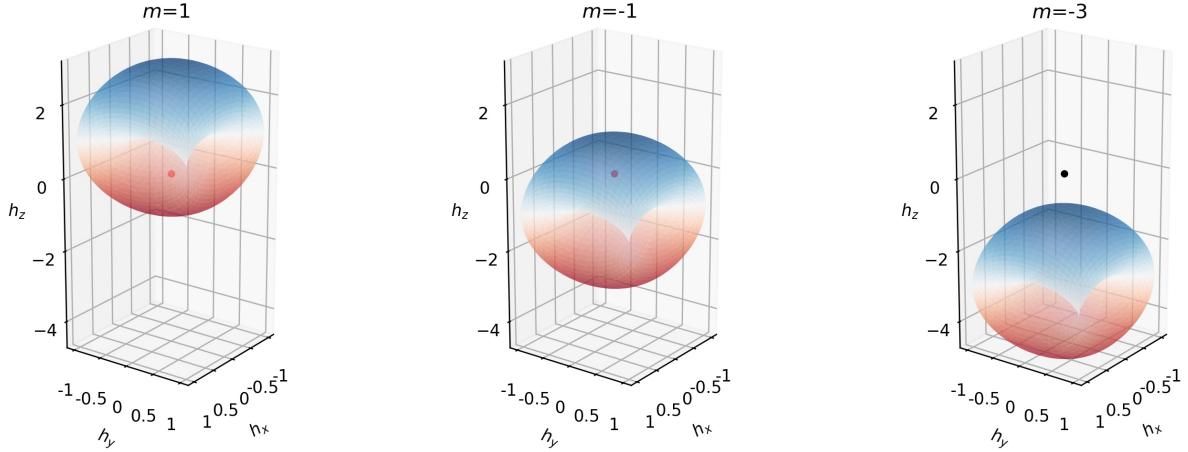


Figure 13: Image of the Brillouin zone under $\mathbf{d}(\mathbf{k})$ for $m = \pm 1, -3$.

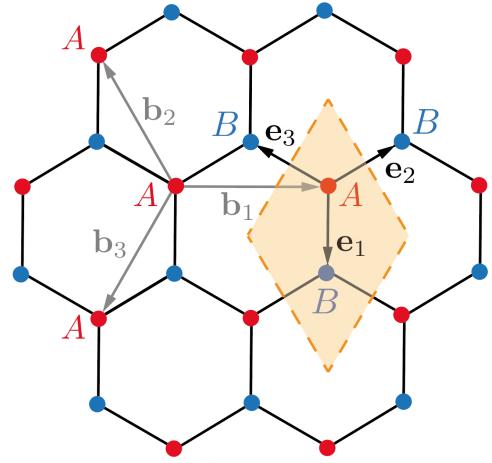


Figure 14: The bipartite honeycomb lattice labelled with letters A and B , with primitive vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ and lattice vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$. The unit cell is shadowed in orange.

spacing between adjacent sites be a , then the unit and lattice vectors are given by

$$\begin{aligned}\mathbf{e}_1 &= -a\mathbf{e}_y & \mathbf{b}_1 &= \mathbf{e}_2 - \mathbf{e}_3 = \sqrt{3}a\mathbf{e}_x \\ \mathbf{e}_2 &= a\left(\frac{\sqrt{3}}{2}\mathbf{e}_x + \frac{1}{2}\mathbf{e}_y\right) & \mathbf{b}_2 &= \mathbf{e}_3 - \mathbf{e}_1 = \sqrt{3}a\left(-\frac{1}{2}\mathbf{e}_x + \frac{\sqrt{3}}{2}\mathbf{e}_y\right) \\ \mathbf{e}_3 &= a\left(-\frac{\sqrt{3}}{2}\mathbf{e}_x + \frac{1}{2}\mathbf{e}_y\right) & \mathbf{b}_3 &= \mathbf{e}_1 - \mathbf{e}_2 = \sqrt{3}a\left(-\frac{1}{2}\mathbf{e}_x - \frac{\sqrt{3}}{2}\mathbf{e}_y\right)\end{aligned}$$

Note that the honeycomb lattice is not a Bravais lattice, but can be considered as two superposed, interlacing triangular (Bravais) sublattices given by joining the A sites together, and by joining the B sites together. This lattice sub-symmetry can be exploited to write tight-binding hamiltonians in momentum space as two-level systems.

Since we will be performing Fourier transforms it is useful to look at the Brillouin zone of the honeycomb

lattice. Using $(\mathbf{b}_2, \mathbf{b}_3)$ as lattice vectors then the reciprocal lattice vectors are given by

$$\mathbf{a}_1 = 2\pi \frac{\mathbf{R}_{\frac{\pi}{2}} \mathbf{b}_3}{\mathbf{b}_2 \cdot \mathbf{R}_{\frac{\pi}{2}} \mathbf{b}_3} = \frac{2\pi}{\sqrt{3}a} \left(-\mathbf{e}_x + \frac{1}{\sqrt{3}}\mathbf{e}_y \right) \quad (6.1.1)$$

$$\mathbf{a}_2 = 2\pi \frac{\mathbf{R}_{\frac{\pi}{2}} \mathbf{b}_2}{\mathbf{b}_3 \cdot \mathbf{R}_{\frac{\pi}{2}} \mathbf{b}_2} = -\frac{2\pi}{\sqrt{3}a} (\mathbf{e}_x + \frac{1}{\sqrt{3}}\mathbf{e}_y) \quad (6.1.2)$$

where $\mathbf{R}_{\frac{\pi}{2}}$ is the $\frac{\pi}{2}$ (anti-clockwise) rotation matrix. The reciprocal lattice will therefore be a hexagonal lattice, so the Brillouin zone, which is the Wigner-Seitz cell of the reciprocal lattice will also be a hexagon. Note that the vertices of the hexagon are only two due to the lattice periodicity, all other vertices can be related to them via the reciprocal vectors. These non-unique vertices are known as **Dirac points** and are given by

$$K = \left(\frac{4\pi}{3\sqrt{3}a}, 0 \right), \quad K' = \left(-\frac{4\pi}{3\sqrt{3}a}, 0 \right) \quad (6.1.3)$$

Since, it will be helpful to exploit the periodicity of the reciprocal lattice to rearrange the FBZ into a rectangle, as shown in fig. 15.

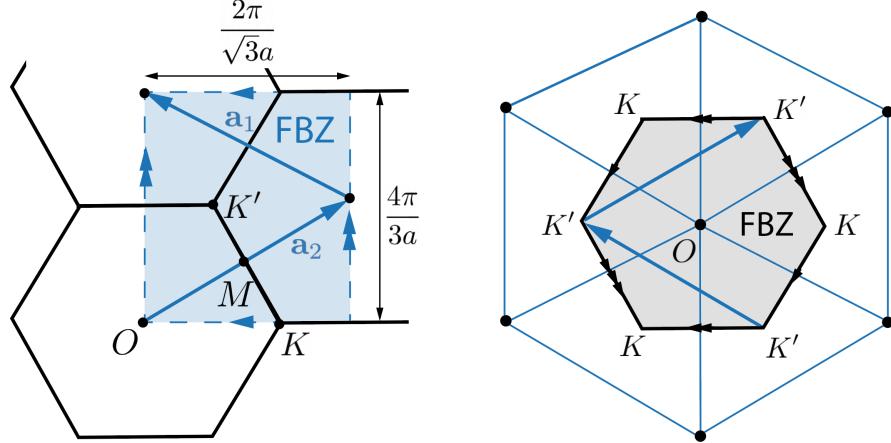


Figure 15: Right: the reciprocal lattice is a hexagonal lattice with lattice vectors $\mathbf{a}_1, \mathbf{a}_2$ shown in blue. The FBZ is a hexagon, shaded in grey. Left: the hexagonal BZ is rearranged into a rectangle with periodic boundary conditions, thus forming a torus.

6.2 The tight-binding approximation for graphene

Let us consider a simple tight-binding model on a honeycomb lattice:

$$H = -t \sum_{\langle ij \rangle} (c_i^\dagger c_j + h.c.) = -t \sum_{\langle ij \rangle} (a_i^\dagger b_j + h.c.) \quad (6.2.1)$$

We introduce the momentum-space creation and annihilation operators

$$a_{\mathbf{k}} = \sum_i e^{i\mathbf{k} \cdot \mathbf{r}_i} a_i \iff a_i = \sum_{\mathbf{k} \in \text{BZ}} e^{-i\mathbf{k} \cdot \mathbf{r}_i} a_{\mathbf{k}} \quad (6.2.2)$$

and similarly for the B sublattice. Substituting this into (6.2.1) we then obtain

$$\sum_{\langle ij \rangle} a_i^\dagger b_j + h.c. = \sum_{\langle ij \rangle} \sum_{\mathbf{k} \in \text{BZ}} \sum_{\mathbf{q} \in \text{BZ}} e^{i\mathbf{k} \cdot \mathbf{r}_i} e^{-i\mathbf{q} \cdot \mathbf{r}_j} a_\mathbf{k}^\dagger b_\mathbf{q} + h.c. \quad (6.2.3)$$

$$= \sum_i \sum_{\mathbf{k} \in \text{BZ}} \sum_{\mathbf{q} \in \text{BZ}} e^{i(\mathbf{k}-\mathbf{q}) \cdot \mathbf{r}_i} (e^{i\mathbf{q} \cdot \mathbf{e}_1} + e^{i\mathbf{q} \cdot \mathbf{e}_2} + e^{i\mathbf{q} \cdot \mathbf{e}_3}) a_\mathbf{k}^\dagger b_\mathbf{q} + h.c. \quad (6.2.4)$$

$$= \sum_{\mathbf{k} \in \text{BZ}} (e^{i\mathbf{k} \cdot \mathbf{e}_1} + e^{i\mathbf{k} \cdot \mathbf{e}_2} + e^{i\mathbf{k} \cdot \mathbf{e}_3}) a_\mathbf{k}^\dagger b_\mathbf{k} + h.c. \quad (6.2.5)$$

Therefore, we can write the Hamiltonian in the following way

$$H = \sum_{\mathbf{k}} \Psi^\dagger(\mathbf{k}) \mathcal{H}(\mathbf{k}) \Psi(\mathbf{k}) \quad (6.2.6)$$

where we introduced the pseudo-spinor $\Psi(\mathbf{k})^\dagger = (a_\mathbf{k}^\dagger \ b_\mathbf{k}^\dagger)$ and the momentum space Hamiltonian

$$\mathcal{H}(\mathbf{k}) = -t \begin{pmatrix} 0 & e^{i\mathbf{k} \cdot \mathbf{e}_1} + e^{i\mathbf{k} \cdot \mathbf{e}_2} + e^{i\mathbf{k} \cdot \mathbf{e}_3} \\ e^{-i\mathbf{k} \cdot \mathbf{e}_1} + e^{-i\mathbf{k} \cdot \mathbf{e}_2} + e^{-i\mathbf{k} \cdot \mathbf{e}_3} & 0 \end{pmatrix} \quad (6.2.7)$$

The bipartite nature of the honeycomb lattice endowed a pseudo-spin to the Hamiltonian! Since (6.2.7) is a 2×2 matrix it can be expressed as a two-level system hamiltonian using Pauli matrices. The result is

$$\mathcal{H}(\mathbf{k}) = -t \sum_{i=1}^3 (\cos(\mathbf{k} \cdot \mathbf{e}_i) \sigma_x - \sin(\mathbf{k} \cdot \mathbf{e}_i) \sigma_y) \quad (6.2.8)$$

The energy dispersion is then given by

$$E(\mathbf{k}) = \pm t \sqrt{\left(\sum_{i=1}^3 \cos(\mathbf{k} \cdot \mathbf{e}_i) \right)^2 + \left(\sum_{i=1}^3 \sin(\mathbf{k} \cdot \mathbf{e}_i) \right)^2} \quad (6.2.9)$$

Since each atom contributes one electron in each sp² orbital, Pauli's exclusion principle implies that only the lower band will be filled at low temperatures. The only low-energy excitations will occur near the top of the Fermi surface near the points where the gap closes:

$$\begin{cases} \cos(ak_y) + 2 \cos\left(\frac{\sqrt{3}a}{2}k_x\right) \cos\left(\frac{a}{2}k_y\right) = 0 \\ -\sin(ak_y) + 2 \cos\left(\frac{\sqrt{3}a}{2}k_x\right) \sin\left(\frac{a}{2}k_y\right) = 0 \end{cases} \quad (6.2.10)$$

It is easy to see that this only occurs at the Dirac points K and K' . To obtain a low-energy effective theory, let's expand (6.2.7) about these points in the continuum limit. Let $\mathbf{k} = \mathbf{K} + \mathbf{q}$ where \mathbf{q} is the small deviation from the Dirac points, then we find that

$$\cos(ak_y) \approx 1 \quad (6.2.11)$$

$$\cos\left(\frac{\sqrt{3}a}{2}k_x + \frac{a}{2}k_y\right) \sim -\frac{\sqrt{3}a}{2}q_x \sin\left(\frac{2\pi}{3}\right) - \frac{a}{2}q_y \sin\left(\frac{2\pi}{3}\right) \quad (6.2.12)$$

$$\cos\left(-\frac{\sqrt{3}a}{2}k_x + \frac{a}{2}k_y\right) \sim -\frac{\sqrt{3}a}{2}q_x \sin\left(\frac{2\pi}{3}\right) + \frac{a}{2}q_y \sin\left(\frac{2\pi}{3}\right) \quad (6.2.13)$$

and

$$-\sin(ak_y) \sim -aq_y \quad (6.2.14)$$

$$\sin\left(\frac{\sqrt{3}a}{2}k_x + \frac{a}{2}k_y\right) \sim \frac{\sqrt{3}a}{2}q_x \cos\left(\frac{2\pi}{3}\right) + \frac{a}{2}q_y \cos\left(\frac{2\pi}{3}\right) \quad (6.2.15)$$

$$\sin\left(-\frac{\sqrt{3}a}{2}k_x + \frac{a}{2}k_y\right) \sim -\frac{\sqrt{3}a}{2}q_x \cos\left(\frac{2\pi}{3}\right) + \frac{a}{2}q_y \cos\left(\frac{2\pi}{3}\right) \quad (6.2.16)$$

so the low energy expansion is

$$\mathcal{H}(\mathbf{q} \pm \mathbf{K}) \approx \hbar v_F (\mathbf{q}_\mp \cdot \boldsymbol{\sigma}) \implies E(\mathbf{q} + \mathbf{K}) \approx \hbar v_F |\mathbf{q}| \quad (6.2.17)$$

where $\mathbf{q}_\mp = (q_x \mp q_y)$ and $v_F = \frac{3at}{2\hbar}$. We have recovered the equation for massless Dirac fermions travelling with speed v_F rather than c .

To summarize, we have learnt that in the tight-binding approximation, graphene is never fully gapped. This comes as a result of inversion symmetry, the A and B sites in each unit cell are identical so if we were to relabel them as $A \leftrightarrow B$ nothing would have changed. To break this symmetry, and introduce a gap, we need to add a staggered mass term. For example, let's add a staggered on-site potential $m\epsilon_i$

$$H = -t \sum_{\langle ij \rangle} (c_i^\dagger c_j + h.c.) + m \sum_i \epsilon_i n_i \quad (6.2.18)$$

which is $+m$ if site i is in the A sublattice, and $-m$ otherwise. Repeating the calculation from above it is easy to see that

$$\mathcal{H}(\mathbf{k}) = -t \sum_{i=1}^3 (\cos(\mathbf{k} \cdot \mathbf{e}_i) \sigma_x - \sin(\mathbf{k} \cdot \mathbf{e}_i) \sigma_y) + m\sigma_z \quad (6.2.19)$$

and thus

$$E(\mathbf{k}) = \pm \sqrt{m^2 + \left(t \sum_{i=1}^3 \cos(\mathbf{k} \cdot \mathbf{e}_i)\right)^2 + \left(t \sum_{i=1}^3 \sin(\mathbf{k} \cdot \mathbf{e}_i)\right)^2} \quad (6.2.20)$$

This time we can have a band gap assuming $m \neq 0$ and the result is a trivial band insulator. The low-energy effective theory is then

$$\mathcal{H}(\mathbf{q} + \mathbf{K}) \approx \hbar v_F (\pm q_x \sigma_x + q_y \sigma_y) + m\sigma_z \quad (6.2.21)$$

Alternatively, we could have instead added a next-nearest neighbor hopping term. However in this case band gaps could close for certain special values of \mathbf{k} . Nevertheless, it is easy to verify that the Chern number for both bands is zero. Indeed this follows immediately from the unbroken time-reversal symmetry.

6.3 Peierls substitution

To break time-reversal symmetry we can insert a magnetic flux in our honeycomb lattice. This will induce complex hopping terms in the Hamiltonian which will hopefully break TRS.

In the absence of a magnetic field, the tight-binding Hamiltonian is

$$H = \frac{\mathbf{p}^2}{2m} + V(\mathbf{r}) \quad (6.3.1)$$

where V is the ion-electron Coulomb interaction. We define the Wannier states

$$\phi_{\mathbf{R}}(\mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} e^{-i\mathbf{k} \cdot \mathbf{R}} \Psi_{\mathbf{k}}(\mathbf{r}) \text{ where } H\Psi_{\mathbf{k}}(\mathbf{r}) = E_{\mathbf{k}} \Psi_{\mathbf{k}}(\mathbf{r}) \quad (6.3.2)$$

Let's now add a magnetic field \mathbf{B} , and choose a potential \mathbf{A} so that the Hamiltonian becomes

$$H = \frac{(\mathbf{p} + e\mathbf{A})^2}{2m} + V(\mathbf{r}) \quad (6.3.3)$$

Therefore let us define the modified Wannier functions

$$\phi'_{\mathbf{R}}(\mathbf{r}) = \chi(\mathbf{r})\phi_{\mathbf{R}}(\mathbf{r}) \quad (6.3.4)$$

where $\chi(\mathbf{r})$ is an unknown function, and substitute into (6.3.3):

$$H\phi'_{\mathbf{R}}(\mathbf{r}) = \left[\frac{(\mathbf{p} + e\mathbf{A})^2}{2m} + V(\mathbf{r}) \right] (\chi(\mathbf{r})\phi_{\mathbf{R}}(\mathbf{r})) \quad (6.3.5)$$

To simplify our calculations we impose that $(-\hbar\mathbf{\nabla} + e\mathbf{A})(\chi(\mathbf{r})\phi_{\mathbf{R}}(\mathbf{r})) = -i\hbar\chi(\mathbf{r})\nabla\phi_{\mathbf{R}}(\mathbf{r})$ which implies

$$-i\hbar\phi_{\mathbf{R}}(\mathbf{r})\nabla\chi + e\mathbf{A}(\mathbf{r})\chi(\mathbf{r})\phi_{\mathbf{R}}(\mathbf{r}) = 0 \quad (6.3.6)$$

$$\implies \chi(\mathbf{r}) = \exp\left(-\frac{ie}{\hbar} \int_{\mathbf{R}}^{\mathbf{r}} \mathbf{A} \cdot d\mathbf{r}'\right) \quad (6.3.7)$$

Therefore we have found that

$$\phi'_{\mathbf{R}}(\mathbf{r}) = \exp\left(-\frac{ie}{\hbar} \int_{\mathbf{R}}^{\mathbf{r}} \mathbf{A} \cdot d\mathbf{r}'\right)\phi_{\mathbf{R}}(\mathbf{r}) \implies H\phi'_{\mathbf{R}}(\mathbf{r}) = \exp\left(-\frac{ie}{\hbar} \int_{\mathbf{R}}^{\mathbf{r}} \mathbf{A} \cdot d\mathbf{r}'\right)H\phi_{\mathbf{R}}(\mathbf{r}) \quad (6.3.8)$$

The phase $\varphi = \frac{e}{\hbar} \int_{\mathbf{R}}^{\mathbf{r}} \mathbf{A} \cdot d\mathbf{r}'$ is known as the **Peierls's phase** (Luttinger, 1951). The hopping matrix is now given by

$$t'_{\mathbf{R}_1, \mathbf{R}_2} = \int (\phi_{\mathbf{R}_2}(\mathbf{r}))^* H\phi_{\mathbf{R}_1}(\mathbf{r}) d\mathbf{r} = \int \exp\left(-\frac{ie}{\hbar} \int_{\mathbf{R}_2}^{\mathbf{R}_1} \mathbf{A} \cdot d\mathbf{r}'\right) e^{-ie\Phi/\hbar} \phi_{\mathbf{R}_1}(\mathbf{r}) d\mathbf{r} \quad (6.3.9)$$

$$\approx \exp\left(-\frac{ie}{\hbar} \int_{\mathbf{R}_2}^{\mathbf{R}_1} \mathbf{A} \cdot d\mathbf{r}'\right) t_{\mathbf{R}_1, \mathbf{R}_2} \quad (6.3.10)$$

where Φ is the magnetic flux through the triangle with vertices at $\mathbf{r}, \mathbf{R}_1, \mathbf{R}_2$. Since the potential \mathbf{A} is assumed to be smooth across the lattice scale length, we can approximate Suppose a particle hops around a closed loop enclosing a magnetic flux Φ , then we find that the total accumulated phase must be

$$\frac{\varphi}{2\pi} = \frac{1}{2\pi} \sum_i \arg(t_i) = -\frac{e\Phi}{2\pi\hbar} = \frac{\Phi}{\Phi_0} \quad (6.3.11)$$

If the lattice is a torus so that $\frac{\Phi}{\Phi_0} = n \in \mathbb{Z}$, then we see that the accumulated phase must always vanish modulo 2π , so we can assume without loss of generality that nearest neighbor hoppings on our lattice wil not accumulate a complex phase.

6.4 The Haldane model

In addition to the simple tight-binding model we add a complex next nearest neighbour (NNN) hopping and a staggered mass term.

$$H = t_1 \sum_{\langle ij \rangle} c_i^\dagger c_j + t_2 \sum_{\langle\langle ij \rangle\rangle} e^{\pm i\varphi} c_i^\dagger c_j + m \sum_i \epsilon_i n_i \quad (6.4.1)$$

Here NNN hoppings between A sites acquire a Peierls's phase of $e^{i\varphi}$ while NNN hoppings between B sites acquire a Peierls's phase of $e^{-i\varphi}$. We can achieve this by an alternating magnetic flux inside each unit cell, two such arrangements are shown in ???. In both cases the nearest neighbour hoppings enclose a zero net flux so

we can set the complex phase to zero. The next nearest neighbor hoppings however do enclose a net magnetic flux, so for simplicity we partition it equally along each leg.

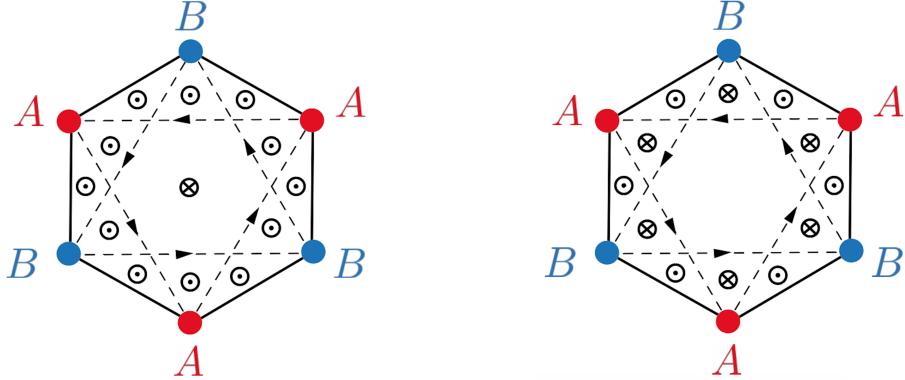


Figure 16: Two possible magnetic flux arrangements for the Haldane model. In both cases the net flux in the hexagonal plaquette is zero, but on each triangular sublattice the enclosed flux is non-zero and endows a phase to the hopping matrix elements.

Once again we wish to take the Fourier transform of (6.4.1). For the next nearest neighbor hopping terms we find that

$$\sum_{\langle\langle ij \rangle\rangle} e^{-i\varphi} a_i^\dagger a_j + e^{i\varphi} b_i^\dagger b_j + h.c. = \sum_{\langle ij \rangle} \sum_{\mathbf{k} \in \text{BZ}} \sum_{\mathbf{q} \in \text{BZ}} e^{-i\mathbf{k} \cdot \mathbf{r}_i} e^{i\mathbf{q} \cdot \mathbf{r}_j} (e^{-i\varphi} a_{\mathbf{k}}^\dagger a_{\mathbf{q}} + e^{i\varphi} b_{\mathbf{k}}^\dagger b_{\mathbf{q}}) + h.c. \quad (6.4.2)$$

$$= \sum_i \sum_{\mathbf{k} \in \text{BZ}} \sum_{\mathbf{q} \in \text{BZ}} e^{i(\mathbf{k}-\mathbf{q}) \cdot \mathbf{r}_i} (e^{i\mathbf{q} \cdot \mathbf{b}_1} + e^{i\mathbf{q} \cdot \mathbf{b}_2} + e^{i\mathbf{q} \cdot \mathbf{b}_3}) (e^{-i\varphi} a_{\mathbf{k}}^\dagger a_{\mathbf{q}} + e^{i\varphi} b_{\mathbf{k}}^\dagger b_{\mathbf{q}}) + h.c. \quad (6.4.3)$$

$$= 2 \sum_{\mathbf{k} \in \text{BZ}} \sum_{i=1}^3 (\cos(\mathbf{k} \cdot \mathbf{b}_i - \varphi) a_{\mathbf{k}}^\dagger a_{\mathbf{k}} + \cos(\mathbf{k} \cdot \mathbf{b}_i + \varphi) b_{\mathbf{k}}^\dagger b_{\mathbf{k}}) \quad (6.4.4)$$

The momentum space Hamiltonian thus reads

$$\mathcal{H}(\mathbf{k}) = \sum_{i=1}^3 \begin{pmatrix} m + 2t_2 \cos(\mathbf{k} \cdot \mathbf{b}_i - \varphi) & t_1 e^{i\mathbf{k} \cdot \mathbf{e}_i} \\ t_1 e^{i\mathbf{k} \cdot \mathbf{e}_i} & -m + 2t_2 \cos(\mathbf{k} \cdot \mathbf{b}_i + \varphi) \end{pmatrix} \quad (6.4.5)$$

Once again we can write this as a Dirac hamiltonian

$$\mathcal{H}(\mathbf{k}) = \epsilon(\mathbf{k}) \mathbb{1} + \mathbf{d}(\mathbf{k}) \cdot \boldsymbol{\sigma} \quad (6.4.6)$$

$$\epsilon(\mathbf{k}) = 2t_2 \cos \varphi \sum_{i=1}^3 \cos(\mathbf{k} \cdot \mathbf{b}_i), \quad \mathbf{d}(\mathbf{k}) = \begin{pmatrix} \sum_{i=1}^3 \cos(\mathbf{k} \cdot \mathbf{e}_i) \\ -\sum_{i=1}^3 \sin(\mathbf{k} \cdot \mathbf{e}_i) \\ m + 2t_2 \sin \varphi \sum_{i=1}^3 \sin(\mathbf{k} \cdot \mathbf{b}_i) \end{pmatrix} \quad (6.4.7)$$

Suppose we start out in the ground state of an insulator. We can adiabatically vary the parameters of our Hamiltonian, and as long as we do not close the band gap we will remain in the ground state of an insulator. We define these two insulators to be topologically equivalent, we can continuously deform their bands into each other.

When the band gap closes the ground states are no longer necessarily adiabatically continuous due to band crossings i.e. adiabatic theorem no longer applies. It is at these points that we should expect the topological properties of our system to vary. Since we are in the low \mathbf{q} limit our result for the Chern number change for

continuous models is applicable.

Following the same argument as for the QWZ model we find the following phase diagram

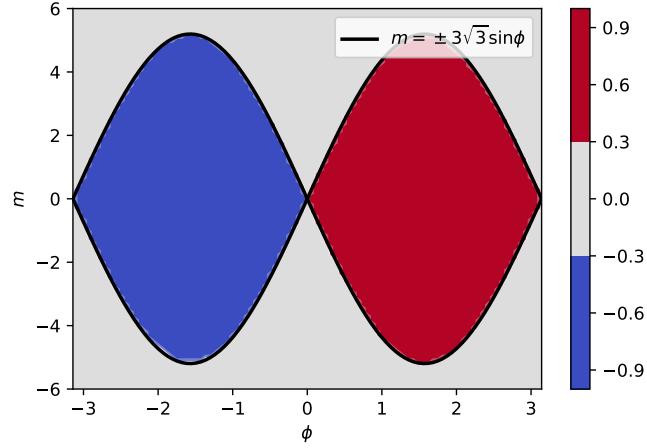


Figure 17: Topological phase diagram of the Haldane model.

6.5 Edge states in the Haldane model

Below we plot the band dispersion for the Haldane model on a finite honeycomb lattice with zig-zag boundary conditions:

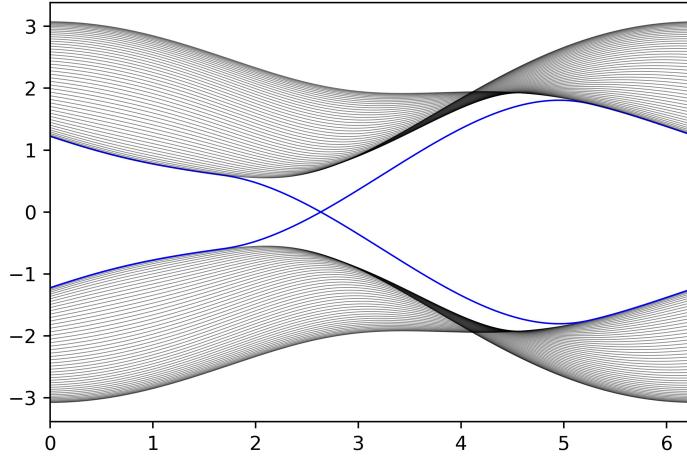


Figure 18: Edge state in the Haldane model are highlighted in blue.

7 Topological insulators and the \mathbb{Z}_2 invariant

7.1 The Kane-Mele model

7.2 The SSH model and polyacetylene

8 Spin systems

We take a break from topological insulators and concern ourselves with an entirely different class of systems, spin models.

The prototypical spin model is the Heisenberg model

$$H = J \sum_{\langle ij \rangle} \mathbf{S}_i \cdot \mathbf{S}_j \quad (8.0.1)$$

Much like with the quantum hall effects, we would like to establish a Lagrangian for the system. While this was relatively straightforward to do with fermionic systems, it is unclear how one should obtain a spin Lagrangian $L = p_i \dot{q}^i - H$ from (8.0.1). What are our canonical coordinates q and p ? In classical field theory this is not too hard, one simply derives the classical equations of motion from the Heisenberg model, and then by some stroke of intuition writes down a Lagrangian which reproduces them.

Our plan is to construct the path integral for the Heisenberg model and read off the Lagrangian from the action. This will yield some interesting (topological) surprises!

8.1 Coherent spin-states

We begin by defining a class of states known as coherent spin-states which will be particularly convenient in deriving the spin path integral. We consider the spin-S states $|S, M\rangle$ which are eigenstates of \mathbf{S}^2 and S_z :

$$\mathbf{S}^2 |S, M\rangle = S(S+1) |S, M\rangle \quad (8.1.1)$$

$$S_z |S, M\rangle = M |S, M\rangle \quad (8.1.2)$$

The spin-coherent state $|\mathbf{n}\rangle$ is given by rotating the maximum weight state $|S, S\rangle$ so as to “point” along \mathbf{n} :

$$|\mathbf{n}\rangle = e^{i\theta(\mathbf{n}_0 \times \mathbf{n}) \cdot \mathbf{S}} |S, S\rangle \quad (8.1.3)$$

It is simple to verify that these states obey

$$\mathbf{n} \cdot \mathbf{S} |\mathbf{n}\rangle = S |\mathbf{n}\rangle \quad (8.1.4)$$

so we may view $|\mathbf{n}\rangle$ is a spin state polarised along \mathbf{n} , it is as classical as a spin-S state can get.

With a bit more effort one can also show that $\{|\mathbf{n}\rangle\}$ is an overcomplete basis (it is complete but not minimal). Here we present a heuristic argument. Firstly let’s consider

$$\int_{S^2} \frac{d^2 \mathbf{n}}{4\pi} (2S+1) |\mathbf{n}\rangle \langle \mathbf{n}| \stackrel{?}{=} \mathbb{1} \quad (8.1.5)$$

where we integrate over all $\mathbf{n} \in S^2$. The RHS is a rotationally invariant operator, so it must be proportional to the identity matrix, in fact it is equal to it (we see this by taking the trace of both sides).

We now see that the advantages of using the coherent states are two-fold:

- (i) they are as classical as spin can get, so they are the “right” objects to use in constructing the Feynman path integral which traditionally takes in a classical Lagrangian.
- (ii) coherent spin states $|\mathbf{n}\rangle$ are defined on a differentiable manifold (the Bloch sphere S^2) which allows us to perform a continuous integral over \mathbf{n} , a crucial ingredient for the path integral recipe.

As a final note, we can view general spin-S coherent states as a tensor product of 2S spin-1/2 coherent states. Indeed by the principles of addition of angular momenta we have

$$|S, S\rangle = \bigotimes_{i=1}^{2S} |\frac{1}{2}, \frac{1}{2}\rangle_i = |\frac{1}{2}, \frac{1}{2}\rangle_1 \otimes |\frac{1}{2}, \frac{1}{2}\rangle_2 \otimes \dots \otimes |\frac{1}{2}, \frac{1}{2}\rangle_{2S} \quad (8.1.6)$$

The spin rotation operator acts individually on each spin-1/2 since $\mathbf{S} = \mathbf{S}_1 + \dots + \mathbf{S}_{2S}$. Thus

$$|\mathbf{n}\rangle = e^{i\theta(\mathbf{n}_0 \times \mathbf{n}) \cdot \mathbf{S}} |S, S\rangle = \left(e^{i\theta(\mathbf{n}_0 \times \mathbf{n}) \cdot \mathbf{S}_1} |\frac{1}{2}, \frac{1}{2}\rangle_1 \right) \otimes \dots \otimes \left(e^{i\theta(\mathbf{n}_0 \times \mathbf{n}) \cdot \mathbf{S}_{2S}} |\frac{1}{2}, \frac{1}{2}\rangle_{2S} \right) = |\mathbf{n}_1\rangle \otimes \dots \otimes |\mathbf{n}_{2S}\rangle \quad (8.1.7)$$

This allows us to prove several properties of spin coherent states by just considering the simpler spin-1/2 case, and then using the tensor product structure of the more general spin-S states.

8.2 Getting the path integral

We are now ready to construct the spin path integral in Euclidean space-time. We find that

$$Z = \text{Tr}(e^{-\beta H}) = (2S+1) \int \frac{d^2\mathbf{n}}{4\pi} \langle \mathbf{n} | e^{-\beta H} | \mathbf{n} \rangle \quad (8.2.1)$$

We split $e^{-\beta H}$ into a product of N terms $e^{-\epsilon H}$ where $N\epsilon = \beta$. Then

$$Z = (2S+1) \int_{S^2} \frac{d^2\mathbf{n}}{4\pi} \langle \mathbf{n} | e^{-\epsilon H} e^{-\epsilon H} \dots e^{-\epsilon H} | \mathbf{n} \rangle \quad (8.2.2)$$

We now insert an identity closure between each exponential term to get

$$Z = (2S+1)^N \int_{\mathbf{n}_{N+1}=\mathbf{n}_1} \left(\prod_{j=1}^N \frac{d^2\mathbf{n}_j}{4\pi} \right) \prod_{j=1}^N \langle \mathbf{n}_{j+1} | e^{-\epsilon H} | \mathbf{n}_j \rangle \quad (8.2.3)$$

Since $N \rightarrow \infty$ we can expand $e^{-\epsilon H} \approx 1 - \epsilon H$. Therefore we find

$$\langle \mathbf{n}_{j+1} | e^{-\epsilon H} | \mathbf{n}_j \rangle = \langle \mathbf{n}_{j+1} | \mathbf{n}_j \rangle \left(1 - \epsilon \frac{\langle \mathbf{n}_{j+1} | H | \mathbf{n}_j \rangle}{\langle \mathbf{n}_{j+1} | \mathbf{n}_j \rangle} \right) \quad (8.2.4)$$

and going to the continuum limit $\epsilon \rightarrow 0$ where $|\mathbf{n}_j\rangle \rightarrow |\mathbf{n}(\tau)\rangle$ then

$$\langle \mathbf{n}_{j+1} | \mathbf{n}_j \rangle \rightarrow \langle \mathbf{n}(\tau + \epsilon) | \mathbf{n}(\tau) \rangle \approx 1 - \epsilon \langle \mathbf{n}(\tau) | \partial_\tau | \mathbf{n}(\tau) \rangle \quad (8.2.5)$$

$$\epsilon \frac{\langle \mathbf{n}_{j+1} | H | \mathbf{n}_j \rangle}{\langle \mathbf{n}_{j+1} | \mathbf{n}_j \rangle} \rightarrow \epsilon \frac{\langle \mathbf{n}(\tau + \epsilon) | H | \mathbf{n}(\tau) \rangle}{\langle \mathbf{n}(\tau + \epsilon) | \mathbf{n}(\tau) \rangle} \approx \epsilon \langle \mathbf{n}(\tau) | H | \mathbf{n}(\tau) \rangle \quad (8.2.6)$$

so

$$\langle \mathbf{n}_{j+1} | e^{-\epsilon H} | \mathbf{n}_j \rangle \approx (1 - \epsilon \langle \mathbf{n}(\tau) | \partial_\tau | \mathbf{n}(\tau) \rangle)(1 - \epsilon \langle \mathbf{n}(\tau) | H | \mathbf{n}(\tau) \rangle) \approx 1 - \epsilon (\langle \mathbf{n}(\tau) | \partial_\tau | \mathbf{n}(\tau) \rangle + \langle \mathbf{n}(\tau) | H | \mathbf{n}(\tau) \rangle) \quad (8.2.7)$$

Had we worked out the higher order terms we would have gotten a power series expansion for an exponential, so we have obtained the following (Euclidean) path integral

$$Z_E = \int_{\mathbf{n}(0)=\mathbf{n}(\beta)} \mathcal{D}[\mathbf{n}] \exp \left(- \int_0^\beta d\tau [\langle \mathbf{n}(\tau) | \partial_\tau | \mathbf{n}(\tau) \rangle + \langle \mathbf{n}(\tau) | H | \mathbf{n}(\tau) \rangle] \right) \quad (8.2.8)$$

where

$$\mathcal{D}[\mathbf{n}] \equiv \lim_{N \rightarrow \infty} \left(\prod_{j=1}^N \frac{d^2 \mathbf{n}_j}{4\pi} (2S+1) \right) \quad (8.2.9)$$

We then read off the Lagrangian

$$\mathcal{L}_E[\mathbf{n}] = \langle \mathbf{n}(\tau) | \partial_\tau | \mathbf{n}(\tau) \rangle + \langle \mathbf{n}(\tau) | H | \mathbf{n}(\tau) \rangle \quad (8.2.10)$$

which in Minkowski space-time reads

$$\mathcal{L}_M[\mathbf{n}] = i \langle \mathbf{n}(t) | \partial_t | \mathbf{n}(t) \rangle - \langle \mathbf{n}(t) | H | \mathbf{n}(t) \rangle \quad (8.2.11)$$

The Weiss-Zumino term

The action we derived contains a Berry-phase term, known as a Wess-Zumino term

$$S_{WZ} = i \int_0^T dt \langle \mathbf{n}(t) | \partial_t | \mathbf{n}(t) \rangle \quad (8.2.12)$$

Let's consider a spin-1/2 particle where we can explicitly construct the spin-coherent states $|\mathbf{n}\rangle$ where $\mathbf{n} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$. We work in the gauge where:

$$|\mathbf{n}\rangle = \begin{pmatrix} e^{-i\phi} \sin \frac{\theta}{2} \\ \sin \frac{\theta}{2} \end{pmatrix} \quad (8.2.13)$$

then

$$\langle \mathbf{n} | \partial_t | \mathbf{n} \rangle = -i\dot{\phi} \sin^2 \frac{\theta}{2} = -\frac{i}{2} \dot{\phi} (1 - \cos \theta) \quad (8.2.14)$$

Since \mathbf{n} lies on the unit sphere we have $\dot{\mathbf{n}} = \dot{\theta} \mathbf{e}_\theta + \dot{\phi} \sin \theta \mathbf{e}_\phi$, and thus $\langle \mathbf{n} | \partial_t | \mathbf{n} \rangle = -\frac{i}{2} \dot{\mathbf{n}} \cdot \mathbf{A}$ where

$$\mathbf{A} = \frac{1 - \cos \theta}{\sin \theta} \mathbf{e}_\phi \quad (8.2.15)$$

Therefore we find that the Berry phase of a free spin-1/2

$$S_{WZ}[\mathbf{n}] = \frac{1}{2} \int_0^T dt \langle \mathbf{n} | \partial_t | \mathbf{n} \rangle = \frac{1}{2} \int_0^T dt \dot{\mathbf{n}} \cdot \mathbf{A} = \frac{1}{2} \int_{\gamma} d\mathbf{n} \cdot \mathbf{A} \quad (8.2.16)$$

and using Stoke's theorem this turns into a surface integral

$$S_{WZ}[\mathbf{n}] = \frac{1}{2} \int_{S_{\gamma}} d^2 \mathbf{n} \cdot \mathbf{B} \quad (8.2.17)$$

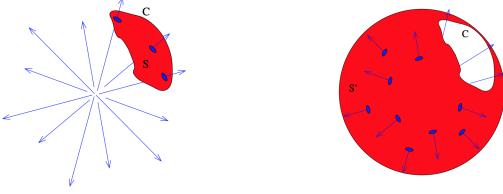
where $\mathbf{B} = \nabla \times \mathbf{A} = \frac{1}{r} \mathbf{e}_r - \frac{1 - \cos \theta}{r \sin \theta} \mathbf{e}_\theta$ is the Berry curvature. Thus we find that the Wess-Zumino term yields the area enclosed by the curve S_{γ} :

$$S_{WZ}[\mathbf{n}] = \frac{1}{2} \int_{S_{\gamma}} d^2 \mathbf{n} = \frac{1}{2} \mathcal{A}[S_{\gamma}] \quad (8.2.18)$$

However this, because the sphere is a closed manifold without a boundary so two possible definitions of S_{γ} are possible. We see that $\mathcal{A}[S_{\gamma_1}] = 4\pi - \mathcal{A}[S_{\gamma_2}]$ which at first appears to be a problem, our action is not well-defined. However any physical quantity is expressed using the exponential of the action, so we need only to worry about the well-definedness of the action modulo 2π .

We see that

$$e^{i\mathcal{A}[S_{\gamma_1}]/2} = e^{i\mathcal{A}[S_{\gamma_2}]/2} \quad (8.2.19)$$



as required, both choices of S_γ yield identical physical results.

We can actually extend our calculation to general spin-S path integrals using the fact that spin-S coherent states are just tensor products of spin-1/2 coherent states. Then

$$S_{WZ}[\mathbf{n}] = S\mathcal{A}[S_\gamma] \quad (8.2.20)$$

Now the condition that $e^{i\mathcal{A}[S_{\gamma_1}]/2} = e^{i\mathcal{A}[S_{\gamma_2}]/2}$ implies that S is either an integer or a half-integer, spin must be quantised!

Note that $S_{WZ}[\mathbf{n}]$ can be expressed in a coordinate invariant way as

$$S_{WZ}[\mathbf{n}] = S \int_0^T dt \int_0^1 d\tau \mathbf{n} \cdot \left(\frac{\partial \mathbf{n}}{\partial t} \times \frac{\partial \mathbf{n}}{\partial \tau} \right) \quad (8.2.21)$$

where we parametrise $\mathbf{n} = \mathbf{n}(t, \tau)$ such that $\mathbf{n}(t, 0) = (0, 0, 1) \equiv \mathbf{n}_0$ and $\mathbf{n}(t, 1) = \mathbf{n}(t)$. The quantity on the RHS of (8.2.21) is just the winding number (times 4π) we encountered previously, in the high energy community it is also referred to as a **Pontryagin index**.

A more formal proof of this equivalence is presented below.

8.3 Weiss-Zumino = Pontryagin

We can write the spin coherent states as

$$|\mathbf{N}\rangle = e^{i\theta(\mathbf{n}_0 \times \mathbf{n}) \cdot \mathbf{S}} |S, S\rangle = e^{zS_+ - z^*S_-} |S, S\rangle \quad (8.3.1)$$

where $z = \frac{1}{2} \sin \theta e^{-i\phi}$. We define the following matrix of spin operators

$$\mathcal{S} = \begin{pmatrix} S_z & S_x - iS_y \\ S_x + iS_y & -S_z \end{pmatrix} \implies \langle \mathbf{N} | \mathcal{S} | \mathbf{N} \rangle = SW \quad (8.3.2)$$

where

$$W = \mathbf{N} \cdot \boldsymbol{\sigma} = \begin{pmatrix} N_z & N_x - iN_y \\ N_x + iN_y & -N_z \end{pmatrix} \quad (8.3.3)$$

For convenience we also define the anti-Hermitian matrix $U = \exp \left[\begin{pmatrix} 0 & z \\ -z^* & 0 \end{pmatrix} \right]$ so that $W = U\sigma_z U^\dagger$. We can use the identity from Linear algebra

$$\frac{d}{dt} (\exp(A(t))) = \int_0^1 du e^{(1-u)A} \frac{dA}{dt} e^{uA} \quad (8.3.4)$$

to find that in our case

$$\frac{d}{dt}|\mathbf{N}(t)\rangle = \frac{d}{dt}(e^{zS_+ - z^*S_-}|S, S\rangle) = \int_0^1 du e^{(1-u)(zS_+ - z^*S_-)} \left(\frac{\partial z}{\partial t}S_+ - \frac{\partial z^*}{\partial t}S_- \right) e^{u(zS_+ - z^*S_-)}|S, S\rangle \quad (8.3.5)$$

$$\implies \langle \mathbf{N} | \partial_t |\mathbf{N}\rangle = \int_0^1 du \langle 0 | e^{-u(zS_+ - z^*S_-)} \left(\frac{\partial z}{\partial t}S_+ - \frac{\partial z^*}{\partial t}S_- \right) e^{u(zS_+ - z^*S_-)} | 0 \rangle \quad (8.3.6)$$

Defining the parametrised coherent states

$$|\mathbf{N}(t, u)\rangle = e^{u(zS_+ - z^*S_-)}|S, S\rangle \quad (8.3.7)$$

then the Berry phase is

$$\int_0^T dt \langle \mathbf{N}(t) | \frac{d}{dt} |\mathbf{N}(t)\rangle = \int_0^T dt \int_0^1 du \langle \mathbf{N}(t, u) | \frac{\partial z}{\partial t}S_+ - \frac{\partial z^*}{\partial t}S_- | \mathbf{N}(t, u) \rangle \quad (8.3.8)$$

$$= S \int_0^T dt \int_0^1 du \left[\frac{\partial z}{\partial t}W_{21}(t, u) - \frac{\partial z^*}{\partial t}W_{12}(t, u) \right] \quad (8.3.9)$$

$$= S \int_0^T dt \int_0^1 du \text{Tr} \left[\begin{pmatrix} 0 & z \\ -z^* & 0 \end{pmatrix} \frac{\partial W(t, u)}{\partial t} \right] \quad (8.3.10)$$

after integrating by parts and noting $z(0) = z(T)$. Here we defined

$$W(t, u) = \mathbf{N}(t, u) \cdot \boldsymbol{\sigma} \quad (8.3.11)$$

which is idempotent and obeys

$$-\frac{1}{2}W(t, u) \frac{\partial W(t, u)}{\partial u} = \begin{pmatrix} 0 & z \\ -z^* & 0 \end{pmatrix} \quad (8.3.12)$$

Consequently

$$\int_0^T dt \langle \mathbf{N}(t, u) | \frac{d}{dt} |\mathbf{N}(t, u) \rangle = \frac{S}{2} \int_0^T dt \int_0^1 du \text{Tr} \left[W \frac{\partial W}{\partial u} \frac{\partial W}{\partial t} \right] \quad (8.3.13)$$

$$= \frac{S}{2} \int_0^T dt \int_0^1 du \text{Tr} \left[(\mathbf{N} \cdot \boldsymbol{\sigma}) \left(\frac{\partial \mathbf{N}}{\partial u} \cdot \boldsymbol{\sigma} \right) \left(\frac{\partial \mathbf{N}}{\partial t} \cdot \boldsymbol{\sigma} \right) \right] \quad (8.3.14)$$

Finally we use the property of Pauli matrices that $\text{Tr}[\mathbf{A} \cdot \boldsymbol{\sigma}] = 0$ and $\text{Tr}[(\mathbf{A} \cdot \boldsymbol{\sigma})(\mathbf{B} \cdot \boldsymbol{\sigma})] = 2\mathbf{A} \cdot \mathbf{B}$ to get

$$\text{Tr}[(\mathbf{A} \cdot \boldsymbol{\sigma})(\mathbf{B} \cdot \boldsymbol{\sigma})(\mathbf{C} \cdot \boldsymbol{\sigma})] = \text{Tr}[(\mathbf{A} \cdot \mathbf{B})(\mathbf{C} \cdot \boldsymbol{\sigma}) + i(\mathbf{C} \cdot \boldsymbol{\sigma})(\mathbf{A} \times \mathbf{B}) \cdot \boldsymbol{\sigma}] \quad (8.3.15)$$

$$= 2i\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) \quad (8.3.16)$$

Using this we finally find that

$$\int_0^T dt \langle \mathbf{N}(t, u) | \frac{d}{dt} |\mathbf{N}(t, u) \rangle = -iS \int_0^T dt \int_0^1 du \mathbf{N} \cdot \left(\frac{\partial \mathbf{N}}{\partial t} \times \frac{\partial \mathbf{N}}{\partial u} \right) \quad (8.3.17)$$

as desired.

8.4 Heisenberg AF chain: θ -terms

We consider a Heisenberg antiferromagnetic chain (with periodic boundary conditions)

$$H = J \sum_i \mathbf{S}_i \cdot \mathbf{S}_{i+1} \quad (8.4.1)$$

where $J > 0$. It is straightforward to generalise our single-spin path integral to many-spin systems using properties of tensor products. We see that

$$\langle \mathbf{n}(t) | H | \mathbf{n}(t) \rangle = J \sum_i \langle \mathbf{n}_i(t) | \mathbf{S}_i | \mathbf{n}_i(t) \rangle \langle \mathbf{n}_{i+1}(t) | \mathbf{S}_{i+1} | \mathbf{n}_{i+1}(t) \rangle = JS^2 \sum_i \mathbf{n}_i(t) \cdot \mathbf{n}_{i+1}(t) \quad (8.4.2)$$

so the action takes the form

$$S[\mathbf{n}] = \sum_i \left(S_{WZ}[\mathbf{n}_i] + \frac{JS^2}{2} \int_0^T dt (\mathbf{n}_i(t) - \mathbf{n}_{i+1}(t))^2 \right) \quad (8.4.3)$$

where we used the fact that $\mathbf{n}_i^2(t) = 1$. Moreover, we expect in the low energy limit to have antiferromagnetic ordering in the fields \mathbf{n}_i so it will be helpful to stagger the spins by introducing the transformation $\mathbf{n}_j(t) \rightarrow (-1)^j \mathbf{n}_j(t)$. The staggered action becomes

$$S[\mathbf{n}] = \sum_i \left((-1)^i S_{WZ}[\mathbf{n}_i] - \frac{JS^2}{2} \int_0^T dt (\mathbf{n}_i(t) - \mathbf{n}_{i+1}(t))^2 \right) \quad (8.4.4)$$

To explore fluctuations around the Neel state we separate \mathbf{n}_i into a slowly varying, large order parameter field \mathbf{m}_i and a rapidly varying, small field $(-1)^i a_0 \mathbf{l}$:

$$\mathbf{n}_i(t) = \mathbf{m}_i(t) + (-1)^i a_0 \mathbf{l}_i \quad (8.4.5)$$

This expansion is supported by the fact that low-energy modes in spin-wave theory are found at crystal momenta $\mathbf{k} = 0, \pm\pi$. We should require $\mathbf{n}_i^2 = \mathbf{m}_i^2 = 1$ so that $\mathbf{m} \cdot \mathbf{l} = 0$. Note that \mathbf{m} governs the antiferromagnetic spin ordering while \mathbf{l} governs ferromagnetic spin ordering, so in this language the Neel state corresponds to $\partial_x \mathbf{m} = \mathbf{l} = 0$. Therefore when we will expand about the Neel state in the low-energy/continuum limit we will have to integrate out \mathbf{l} in order to obtain an effective Lagrangian for \mathbf{m} . By this integration of the ferromagnetic excitations we will see that

$$\mathbf{l} \sim \mathbf{m} \times \frac{\partial \mathbf{m}}{\partial t} \quad (8.4.6)$$

Knowing this we can simplify our calculations by *a priori* ignoring any derivatives of \mathbf{l} which will end up becoming second order derivatives of \mathbf{m} .

To go to the continuum limit we set the lattice space constant to be $a_0 \rightarrow 0$ and redefine $\mathbf{n}_i(t) \rightarrow \mathbf{n}(x, t)$. As we just argued, we will work to second order in \mathbf{m}, \mathbf{l} and to first order derivatives in \mathbf{m} and zeroth order derivatives in \mathbf{l} by assuming \mathbf{m}, \mathbf{l} are slowly varying. Then

$$\mathbf{n}(x + a_0, t) - \mathbf{n}(x, t) \approx a_0 \partial_x \mathbf{m}(x, t) + 2a_0 \mathbf{l}(x, t) \quad (8.4.7)$$

and so

$$(\mathbf{n}(r + a_0, t) - \mathbf{n}(r, t))^2 \approx a_0 (\partial_x \mathbf{m})^2 + 4a_0 \mathbf{l}^2 \quad (8.4.8)$$

The Weiss-Zumino term should be treated with more care. We firstly write

$$\sum_i (-1)^i S_{WZ}[\mathbf{n}_i] = \sum_{i \text{ even}} S_{WZ}[\mathbf{n}_i] - \sum_{i \text{ odd}} S_{WZ}[\mathbf{n}_i] \quad (8.4.9)$$

and then perform the usual gradient expansion

$$S_{WZ}[\mathbf{n}_2] - S_{WZ}[\mathbf{n}_1] = S \int_0^T dt \int_0^1 d\tau \left[\mathbf{n}_2 \cdot \left(\frac{\partial \mathbf{n}_2}{\partial t} \times \frac{\partial \mathbf{n}_2}{\partial \tau} \right) - \mathbf{n}_1 \cdot \left(\frac{\partial \mathbf{n}_1}{\partial t} \times \frac{\partial \mathbf{n}_1}{\partial \tau} \right) \right] \quad (8.4.10)$$

$$\rightarrow S \int_0^T dt \int_0^1 d\tau (a_0 \partial_x \mathbf{m} + 2a_0 \mathbf{l}) \cdot \left(\frac{\partial \mathbf{m}}{\partial t} \times \frac{\partial \mathbf{m}}{\partial \tau} \right) \quad (8.4.11)$$

Replacing a sum over sites as a continuous integral $\sum_i \rightarrow \int \frac{dx}{a_0}$ then we obtain

$$S_M[\mathbf{n}] = \int dx \int_0^T dt \left[\frac{S}{2} \mathbf{m} \cdot \left(\frac{\partial \mathbf{m}}{\partial t} \times \frac{\partial \mathbf{m}}{\partial x} \right) - \frac{a_0 JS^2}{2} \left(\frac{\partial \mathbf{m}}{\partial x} \right)^2 - 2a_0 JS^2 \mathbf{l}^2 + S \mathbf{l} \cdot \left(\mathbf{m} \times \frac{\partial \mathbf{m}}{\partial t} \right) \right] \quad (8.4.12)$$

We can integrate out \mathbf{l} by substituting its equations of motion

$$4a_0 JS^2 \mathbf{l} = \mathbf{m} \times \frac{\partial \mathbf{m}}{\partial t} \quad (8.4.13)$$

and finally find that the effective, low-energy Lagrangian of the Heisenberg AF chain is given by

$$\mathcal{L}_M = \frac{1}{2g} \left[\frac{1}{v_s^2} \left(\frac{\partial \mathbf{m}}{\partial t} \right)^2 - \left(\frac{\partial \mathbf{m}}{\partial x} \right)^2 \right] + \frac{\theta}{4\pi} \mathbf{m} \cdot \left(\frac{\partial \mathbf{m}}{\partial t} \times \frac{\partial \mathbf{m}}{\partial x} \right) \quad (8.4.14)$$

where $g = \frac{1}{a_0 JS^2}$, $v_s = 2a_0 JS$ and $\theta = 2\pi S$ are the coupling strength, spin wave velocity and topological angle respectively. The first term is a non-linear O(3) sigma model while the second term, arising from the Berry phase, is a topological term. In Euclidean space-time we can write the action of the this term

$$S_E^\theta = \frac{i\theta}{4\pi} \int dx \int d\tau \mathbf{m} \cdot \left(\frac{\partial \mathbf{m}}{\partial \tau} \times \frac{\partial \mathbf{m}}{\partial x} \right) \equiv i\theta Q \quad (8.4.15)$$

which is known as the topological θ -term. For S_E^θ to be finite we need $\mathbf{m} \rightarrow \mathbf{m}_0$ as $x, \tau \rightarrow \infty$. This conditions implies that our Euclidean space-time is isomorphic to a sphere (i.e. it is a compact manifold), since the field takes the same value at space-time infinity. The action can then be viewed as a winding number of $\mathbf{m}(x, \tau)$ about the Bloch sphere as (x, τ) roams Euclidean space-time, which for smooth fields is necessarily an integer (known as the Pontryagin index or topological charge) so that $Q \in \mathbb{Z}$. This follows from the fact that a continuous mapping \mathbf{m} maps compact manifolds to compact manifolds.

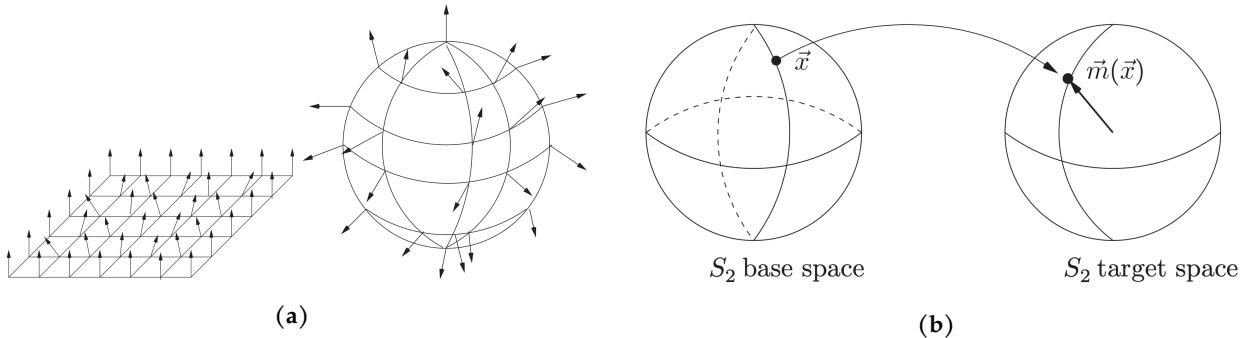
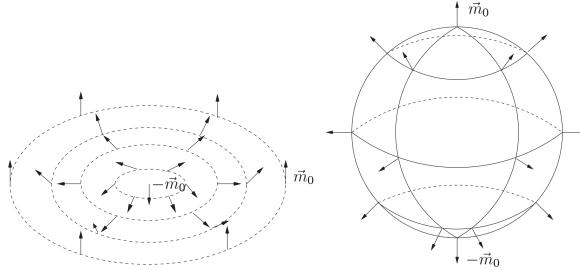


Figure 19: (a)

It will be illustrative to plug in something for \mathbf{m} . Lets take a skyrmion field

$$\mathbf{m} = \frac{1}{1+x^2+y^2}(2x, 2y, x^2 + y^2 - 1) \quad (8.4.16)$$

The field is shown below, it starts out pointing down and slowly curves upwards. By an isomorphism we can embed the field on a sphere, with space-time infinity lying on the north pole, and the origin lying on the south pole. It is straightforward to check upon substitution that $Q = 1$ or this spin configuration. Geometrically this corresponds to there being one magnetic monopole at the origin. Therefore we have found that different



field configurations \mathbf{m} can be classified according to their topological charge. Sectors with topologically distinct charges should exhibit interference effects. In particular we have $e^{-i\theta Q} = (-1)^{2SQ}$ which is $+1$ for integer spin chains and $(-1)^Q$ for half-integer spin chains. Thus different topological sectors interfere only for half-integer spin, while for integer spin chain the topological terms are irrelevant and the action is governed by the kinetic terms. Since the NLσM has a gapped ground state, this observation led Haldane to conjecture that integer spin chains behave much like a NLσM and are thus gapped, while half-integer spin chains much like the spin-1/2 case, are gapless.

Let us now suppose that our base manifold of Euclidean space-time is not compact. This could be done by taking a finite (but long) chain with open boundary conditions, so that a smooth mapping to a sphere is no longer possible. Then Q is no longer forced to be an integer. Note that our derivation of the action did assume periodic boundary conditions, but it can be trivially extended to allow open chains as well.

We have obtained half-integer spin excitations at the ends of an integer-spin antiferromagnet, how bizarre! This goes totally against traditional view that adding integer angular momentum states can only yield other integer angular momentum states.

9 Field theory of topological insulators

We now review the main concepts we have covered in the theory of topological insulators using quantum field theory.

9.1 The parity anomaly

Let's consider a theory of fermions minimally coupled to an external $U(1)$ gauge field:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi \quad (9.1.1)$$

where $D_\mu = \partial_\mu - ieA_\mu$ and $\gamma_\mu = (\sigma_z, i\sigma_x, i\sigma_y)$. We wish to evaluate the linear response of this system and extract the Hall conductivity. To do so we note that (9.1.1) has a $U(1)$ global symmetry $\psi \mapsto e^{i\alpha}\psi$ and $\bar{\psi} \mapsto e^{-i\alpha}\bar{\psi}$ which by Noether's theorem leads to a conserved current

$$j^\mu = \bar{\psi}\gamma^\mu\psi \quad (9.1.2)$$

Consequently we have that

$$\langle j^\mu \rangle = \left\langle \frac{\delta \log \mathcal{Z}}{\delta A_\mu} \right\rangle \quad (9.1.3)$$

where the partition function is

$$\mathcal{Z} = \int \mathcal{D}[\bar{\psi}, \psi] e^{-i \int d^3x \bar{\psi}(i\gamma^\mu D_\mu - m)\psi} \quad (9.1.4)$$

By formal functional integration we find that

$$\log \mathcal{Z} = \log \det(i\gamma^\mu D_\mu - m) = \text{tr} \log(i\gamma^\mu D_\mu - m) \quad (9.1.5)$$

We introduce the Dirac propagator $G_0(x - y)$ satisfying

$$(i\gamma^\mu \partial_\mu - m)G_0(x - y) = \delta(x - y)\mathbb{1} \quad (9.1.6)$$

so that

$$\log \mathcal{Z} = \text{tr} \log(G_0^{-1} + e\gamma^\mu A_\mu) \quad (9.1.7)$$

We now expand the log perturbatively in A finding

$$\log \mathcal{Z} = \text{tr} \log \mathcal{Z}_0 + \text{etr} \underbrace{(G_0 \gamma^\mu A_\mu)}_0 - \frac{1}{2} e^2 \text{tr}(G_0 \gamma^\mu A_\mu G_0 \gamma^\nu A_\nu) + o(A^3) \quad (9.1.8)$$

where the linear term vanishes when traced over. We now evaluate the trace in momentum space:

$$\text{tr}(G_0 \gamma^\mu A_\mu G_0 \gamma^\nu A_\nu) = \sum_q \int d^3x e^{iq \cdot x} G_0(x)(\gamma^\mu A_\mu(x))G_0(x)(\gamma^\nu A_\nu(x))e^{-iq \cdot x} \quad (9.1.9)$$

$$= \sum_{q k_1 k_2} \int d^3x e^{iq \cdot x} \frac{1}{i\gamma^\mu \partial_\mu - m} (\gamma^\mu A_\mu(k_1)) e^{-ik_1 \cdot x} \frac{1}{i\gamma^\mu \partial_\mu - m} (\gamma^\nu A_\nu(k_2)) e^{-ik_2 \cdot x} e^{-iq \cdot x} \quad (9.1.10)$$

$$= \sum_{q k_1 k_2} \int d^3x e^{-i(k_1 + k_2) \cdot x} \frac{1}{i\gamma^\mu (k_1 + k_2 + q)_\mu - m} (\gamma^\mu A_\mu(k_1)) \frac{1}{i\gamma^\mu (k_2 + q)_\mu - m} (\gamma^\nu A_\nu(k_2)) \quad (9.1.11)$$

$$= \sum_{q k} G_0(q)(\gamma^\mu A_\mu(k))G_0(q - k)(\gamma^\nu A_\nu(-k)) \quad (9.1.12)$$

Therefore

$$\log \mathcal{Z} = \sum_k A_\mu(k) \Pi^{\mu\nu}(k) A_\nu(-k) \quad (9.1.13)$$

where we introduced the polarisation tensor $\Pi^{\mu\nu}(k)$ defined as:

$$\Pi^{\mu\nu}(k) = -\frac{1}{2} e^2 \sum_q G_0(k + q) \gamma^\mu G_0(q) \gamma^\nu \quad (9.1.14)$$

The polarisation tensor can be evaluated using Pauli-Villars regularisation, and in the low-momentum, $m \rightarrow \infty$ limit we find that

$$\Pi^{\mu\nu}(k) \approx -\frac{e^2}{4\pi} \text{sgn}(m) \epsilon^{\mu\nu\sigma} k_\sigma \quad (9.1.15)$$

so that

$$\log \mathcal{Z} = \frac{e^2}{4\pi} \text{sgn}(m) \sum_k \epsilon^{\mu\nu\sigma} A_\mu(k) (-k_\sigma) A_\nu(-k) \quad (9.1.16)$$

Hence moving to real space:

$$S_{\text{eff}} = \log \mathcal{Z} = \frac{e^2}{4\pi} \text{sgn}(m) \int d^3x \epsilon^{\mu\nu\sigma} A_\mu \partial_\nu A_\sigma \quad (9.1.17)$$

which induces an effective Chern-Simons term in the action. Consequently we find to first order that

$$\langle j_\mu \rangle = \frac{e}{4\pi} \text{sgn}(m) \epsilon^{\mu\nu\sigma} \partial_\nu A_\sigma \implies \sigma_{xy} = \frac{e}{4\pi} \text{sgn}(m) \quad (9.1.18)$$

This Hall conductivity matches the value we derived previously: each band degeneracy/Dirac point contributes $\frac{e}{4\pi} \text{sgn}(m)$ to the Hall conductivity.

9.2 Generalised first Chern number

We can prove this result in more generality. Due to the nature of the minimal coupling procedure $\partial_\mu \mapsto \partial_\mu - ieA_\mu$ we see that

$$\Pi^{\mu\nu}(k) = \frac{1}{2} \int \frac{d^3q}{(2\pi)^3} \text{Tr} \left[G_0(k+q) \frac{\partial G_0^{-1}(q)}{\partial q_\mu} G_0(q) \frac{\partial G_0^{-1}(q)}{\partial q_\nu} \right] \quad (9.2.1)$$

The Hall conductivity is given by the linear, anti-symmetric part of the polarisation tensor, given by

$$\frac{1}{6} \epsilon_{\mu\nu\sigma} \frac{\partial}{\partial k_\sigma} \Pi^{\mu\nu}(k) \Big|_{k=0} = \frac{e^2}{12} \int \frac{d^3q}{(2\pi)^3} \text{Tr} \left[\frac{\partial G_0(k+q)}{\partial k_\sigma} \Big|_{k=0} \frac{\partial G_0^{-1}(q)}{\partial q_\mu} G_0(q) \frac{\partial G_0^{-1}(q)}{\partial q_\nu} \right] \quad (9.2.2)$$

Using $dG = G(dG^{-1})G$ we finally get

$$\sigma_{xy} = \frac{e}{4\pi} N_2 \quad (9.2.3)$$

where

$$N_2 = \frac{1}{24\pi^3} \int d^3p \epsilon^{\mu\nu\rho} \text{Tr} \left[G_0 \frac{\partial G_0^{-1}}{\partial p_\mu} G_0 \frac{\partial G_0^{-1}}{\partial p_\nu} G_0 \frac{\partial G_0^{-1}}{\partial p_\rho} \right] \quad (9.2.4)$$

is the generalised Chern number. This new expression for the Chern number using Green's functions is applicable even in interacting systems making it a more powerful definition compared to the one using the Berry curvature.

9.3 Nielsen-Ninomiya theorem and chiral edge states

10 The Fractional Quantum Hall effect

10.1 Laughlin's wavefunction

10.2 The Chern-Simons-Landau-Ginzburg action

10.3 Chern-Simons theory

10.4 Effective-low energy theory

10.5 Fractional/anyonic excitations

10.6 Edge excitations

Bibliography

[1] M. V. Berry, Geometric phases in physics , 7 (1989).

[2] M. Suddards, A. Baumgartner, M. Henini, and C. J. Mellor, New Journal of Physics **14**, 083015 (2012).

- [3] D. Tong, “Lectures on the quantum hall effect,” (2016), arXiv:1606.06687 [hep-th] .
- [4] M. O. Goerbig, “Quantum hall effects,” (2009), arXiv:0909.1998 [cond-mat.mes-hall] .
- [5] F. D. M. Haldane, Phys. Rev. Lett. **61**, 2015 (1988).
- [6] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 226801 (2005).
- [7] J. M. Luttinger, Phys. Rev. **84**, 814 (1951).
- [8] S. M. Girvin and K. Yang, *Modern condensed matter physics* (Cambridge University Press, 2019).
- [9] K. v. Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. **45**, 494 (1980).
- [10] K. von Klitzing, Rev. Mod. Phys. **58**, 519 (1986).
- [11] X.-L. Qi and S.-C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011).
- [12] R. Shankar, “Topological insulators – a review,” (2018), arXiv:1804.06471 [cond-mat.str-el] .
- [13] D. Xiao, M.-C. Chang, and Q. Niu, Reviews of modern physics **82**, 1959 (2010).
- [14] J. Cayssol, Comptes Rendus Physique **14**, 760 (2013).
- [15] J. Atteia, *Topology and electronic transport in Dirac systems under irradiation*, Ph.D. thesis, Université de Bordeaux (2018).
- [16] S.-Q. Shen, *Topological insulators*, Vol. 174 (Springer, 2012).
- [17] B. A. Bernevig, *Topological insulators and topological superconductors* (Princeton university press, 2013).
- [18] C. L. Kane, in *Contemporary Concepts of Condensed Matter Science*, Vol. 6 (Elsevier, 2013) pp. 3–34.
- [19] D. Vanderbilt, *Berry Phases in Electronic Structure Theory: Electric Polarization, Orbital Magnetization and Topological Insulators* (Cambridge University Press, 2018).
- [20] D. R. Hofstadter, Phys. Rev. B **14**, 2239 (1976).
- [21] N. W. Ashcroft, N. D. Mermin, *et al.*, “Solid state physics,” (1976).
- [22] R. Peierls, Zeitschrift für Physik **80**, 763 (1933).
- [23] E. Fradkin, *Field theories of condensed matter physics* (Cambridge University Press, 2013).
- [24] A. Altland and B. D. Simons, *Condensed matter field theory* (Cambridge university press, 2010).
- [25] E. Fradkin and M. Stone, Physical Review B **38**, 7215 (1988).
- [26] X.-G. Wen, *Quantum field theory of many-body systems: from the origin of sound to an origin of light and electrons* (OUP Oxford, 2004).
- [27] A. Abanov, in *Topology and Condensed Matter Physics* (Springer, 2017) pp. 281–331.
- [28] S. Sachdev, Physics world **12**, 33 (1999).
- [29] F. D. M. Haldane, Reviews of Modern Physics **89**, 040502 (2017).
- [30] I. Affleck, Journal of Physics: Condensed Matter **1**, 3047 (1989).