5. Chern-Simons Theories

So far we've approached the quantum Hall states from a microscopic perspective, looking at the wavefunctions which describe individual electrons. In this section, we take a step back and describe the quantum Hall effect on a more coarse-grained level. Our goal is to construct effective field theories which capture the response of the quantum Hall ground state to low-energy perturbations. These effective theories are known as *Chern-Simons theories*[45]. They have many interesting properties and, in addition to their role in the quantum Hall effect, play a starring role in several other stories.

Throughout this section, we'll make very general statements about the kind of low-energy effective behaviour that is possible, with very little input about the microscopic properties of the model. As we will see, we will be able to reconstruct many of the phenomena that we've met in the previous chapters.

We will treat the gauge potential A_μ of electromagnetism as a background gauge field. This means that A_μ is not dynamical; it is only a parameter of the theory which tells us which electric and magnetic fields we've turned on. Further, we will not include in A_μ the original background magnetic field which gave rise the Hall effect to begin with. Instead, A_μ will describe only perturbations around a given Hall state, either by turning on an electric field, or by perturbing the applied magnetic field but keeping the kind of state (i.e. the filling fraction) fixed.

In the field theory context, A_μ always couples to the dynamical degrees of freedom through the appropriate current J_μ, so that the action includes the term

$$S_A = \int d^3x\ J^\mu A_\mu \tag{5.1}$$

This is the field theoretic version of (2.8). Note that the measure $\int d^3x$ means that we've assumed that the current lives in a $d = 2+1$ dimensional slice of spacetime; it couples to the gauge field A_μ evaluated on that slice. The action S_A is invariant under gauge transformations $A_\mu \to A_\mu + \partial_\mu \omega$ on account of the conservation of the current

$$\partial_\mu J^\mu = 0$$

These two simple equations will be our starting point for writing down effective field theories that tell us how the system responds when we perturb it by turning on a background electric or magnetic field.

[45]Two reviews on the Chern-Simons approach to the quantum Hall effect are Tony Zee, "*Quantum Hall Fluids*", cond-mat/9501022, Xiao-Gang Wen "*Topological Orders and Edge Excitations in FQH States*", cond-mat/9506066. The discussion here follows the spirit of the beautiful lectures by Edward Witten, "*Three Lectures on Topological Phases of Matter*", arXiv:1510.07698.

5.1 The Integer Quantum Hall Effect

We start by looking at the integer quantum Hall effect. We will say nothing about electrons or Landau levels or anything microscopic. Instead, in our attempt to talk with some generality, we will make just one, seemingly mild, assumption: at low-energies, there are no degrees of freedom that can affect the physics when the system is perturbed.

Let's think about what this assumption means. The first, and most obvious, requirement is that there is a gap to the first excited state. In other words, our system is an insulator rather than a conductor. We're then interested in the physics at energies below this gap.

Naively, you might think that this is enough to ensure that there are no relevant low-energy degrees of freedom. However, there's also a more subtle requirement hiding in our assumption. This is related to the existence of so-called "topological degrees of freedom". We will ignore this subtlety for now, but return to it in Section 5.2 when we discuss the fractional quantum Hall effect.

As usual in quantum field theory, we want to compute the partition function. This is not a function of the dynamical degrees of freedom since these are what we integrate over. Instead, it's a function of the sources which, for us, is the electromagnetic potential A_μ. We write the partition function schematically as

$$Z[A_\mu] = \int D(\text{fields})\ e^{iS[\text{fields};A]/\hbar} \tag{5.2}$$

where "fields" refer to all dynamical degrees of freedom. The action S could be anything at all, as long as it satisfies our assumption above and includes the coupling to A_μ through the current (5.1). We now want to integrate out all these degrees of freedom, to leave ourselves with a theory of the ground state which we write as

$$Z[A_\mu] = e^{iS_{\text{eff}}[A_\mu]/\hbar} \tag{5.3}$$

Our goal is to compute $S_{\text{eff}}[A_\mu]$, which is usually referred to as the *effective action*. Note, however, that it's not the kind of action you meet in classical mechanics. It depends on the parameters of the problem rather than dynamical fields. We don't use it to compute Euler-Lagrange equations since there's no dynamics in A_μ. Nonetheless, it does contain important information since, from the coupling (5.1), we have

$$\frac{\delta S_{\text{eff}}[A]}{\delta A_\mu(x)} = \langle J^\mu(x)\rangle \tag{5.4}$$

This is telling us that the effective action encodes the response of the current to electric and magnetic fields.

Since we don't know what the microscopic Lagrangian is, we can't explicitly do the path integral in (5.2). Instead, our strategy is just to write down all possible terms that can arise and then focus on the most important ones. Thankfully, there are many restrictions on what the answer can be which means that there are just a handful of terms we need to consider. The first restrictions is that the effective action $S_{\rm eff}[A]$ must be gauge invariant. One simple way to achieve this is to construct it out of electric and magnetic fields,

$$\mathbf{E} = -\frac{1}{c}\nabla A_0 - \frac{\partial \mathbf{A}}{\partial t} \quad \text{and} \quad \mathbf{B} = \nabla \times \mathbf{A}$$

The kinds of terms that we can write down are then further restricted by other symmetries that our system may (or may not) have, such as rotational invariance and translational invariance.

Finally, if we care only about long distances, the effective action should be a local functional, meaning that we can write is as $S_{\rm eff}[A] = \int d^dx \ldots$. This property is extremely restrictive. It holds because we're working with a theory with a gap ΔE in the spectrum. The non-locality will only arise at distances comparable to $\sim v\hbar/\Delta E$ with v a characteristic velocity. (This is perhaps most familiar for relativistic theories where the appropriate scale is the Compton wavelength $\hbar/mc$). To ensure that the gap isn't breached, we should also restrict to suitably small electric and magnetic fields.

Now we just have to write down all terms in the effective action that satisfy the above requirements. There's still an infinite number of them but there's a simple organising principle. Because we're interested in small electric and magnetic fields, which vary only over long distances, the most important terms will be those with the fewest powers of A and the fewest derivatives. Our goal is simply to write them down.

Let's first see what all of this means in the context of $d = 3+1$ dimensions. If we have rotational invariance then we can't write down any terms linear in $\mathbf{E}$ or $\mathbf{B}$. The first terms that we can write down are instead

$$S_{\rm eff}[A] = \int d^4x \; \epsilon\mathbf{E}\cdot\mathbf{E} - \frac{1}{\mu}\mathbf{B}\cdot\mathbf{B} \tag{5.5}$$

There is also the possibility of adding a $\mathbf{E}\cdot\mathbf{B}$ term although, when written in terms of A_i this is a total derivative and so doesn't contribute to the response. (This argument is a little bit glib; famously the $\mathbf{E}\cdot\mathbf{B}$ term plays an important role in the subject of 3d topological insulators but this is beyond the scope of these lectures.) The response (5.4) that follows from this effective action is essentially that of free currents. Indeed, it only differs from the familiar Lorentz invariant Maxwell action by the susceptibilities

ϵ and μ which are the free parameters characterising the response of the system. (Note that the response captured by (5.5) isn't quite the same as Ohm's law that we met in Section 1 as there's no dissipation in our current framework).

The action (5.5) has no Hall conductivity because this is ruled out in $d = 3 + 1$ dimensions on rotational grounds. But, as we have seen in great detail, a Hall conductivity is certainly possible in $d = 2 + 1$ dimensions. This means that there must be another kind of term that we can write in the effective action. And indeed there is....

5.1.1 The Chern-Simons Term

The thing that's special in $d = 2+1$ dimension is the existence of the epsilon symbol $\epsilon_{\mu\nu\rho}$ with $\mu, \nu, \rho = 0, 1, 2$. We can then write down a new term, consistent with rotational invariance. The resulting effective action is $S_{\text{eff}}[A] = S_{CS}[A]$ where

$$S_{CS}[A] = \frac{k}{4\pi} \int d^3x \; \epsilon^{\mu\nu\rho} A_\mu \partial_\nu A_\rho \tag{5.6}$$

This is the famous Chern-Simons term. The coefficient k is sometimes called the *level* of the Chern-Simons term.

At first glance, it's not obvious that the Chern-Simons term is gauge invariant since it depends explicitly on A_μ. However, under a gauge transformation, $A_\mu \to A_\mu + \partial_\mu \omega$, we have

$$S_{CS}[A] \to S_{CS}[A] + \frac{k}{4\pi} \int d^3x \; \partial_\mu \left(\omega \epsilon^{\mu\nu\rho} \partial_\nu A_\rho \right)$$

The change is a total derivative. In many situations we can simply throw this total derivative away and the Chern-Simons term is gauge invariant. However, there are some situations where the total derivative does not vanish. Here we will have to think a little harder about what additional restrictions are necessary to ensure that $S_{CS}[A]$ is gauge invariant. We see that the Chern-Simons term is flirting with danger. It's very close to failing the demands of gauge invariance and so being disallowed. The interesting and subtle ways on which it succeeds in retaining gauge invariance will lead to much of the interesting physics.

The Chern-Simons term (5.6) respects rotational invariance, but breaks both parity and time reversal. Here we focus on parity which, in $d = 2 + 1$ dimensions, is defined as

$$x^0 \to x^0 \quad , \quad x^1 \to -x^1 \quad , \quad x^2 \to x^2$$

and, correspondingly, $A_0 \to A_0$, $A_1 \to -A_1$ and $A_2 \to A_2$. The measure $\int d^3x$ is invariant under parity (recall that although $x_1 \to -x_1$, the limits of the integral also change). However, the integrand is not invariant: $\epsilon^{\mu\nu\rho}A_\mu\partial_\nu A_\rho \to -\epsilon^{\mu\nu\rho}A_\mu\partial_\nu A_\rho$. This means that the Chern-Simons effective action with $k \neq 0$ can only arise in systems that break parity. Looking back at the kinds of systems we met in Section 2 which exhibit a Hall conductivity, we see that they all break parity, typically because of a background magnetic field.

Let's look at the physics captured by the Chern-Simons term using (5.4). First, we can compute the current that arises from Chern-Simons term. It is

$$J_i = \frac{\delta S_{CS}[A]}{\delta A_i} = -\frac{k}{2\pi}\epsilon_{ij}E_i$$

In other words, the Chern-Simons action describes a Hall conductivity with

$$\sigma_{xy} = \frac{k}{2\pi} \tag{5.7}$$

This coincides with the Hall conductivity of ν filled Landau levels if we identify the Chern-Simons level with $k = e^2\nu/\hbar$.

We can also compute the charge density J^0. This is given by

$$J_0 = \frac{\delta S_{CS}[A]}{\delta A_0} = \frac{k}{2\pi}B \tag{5.8}$$

Recall that we should think of A_μ as the additional gauge field over and above the original magnetic field. Correspondingly, we should think of J^0 here as the change in the charge density over and above that already present in the ground state. Once again, if we identify $k = e^2\nu/\hbar$ then this is precisely the result we get had we kept ν Landau levels filled while varying $B(\mathbf{x})$.

We see that the Chern-Simons term captures the basic physics of the integer quantum Hall effect, but only if we identify the level $k = e^2\nu/\hbar$. But this is very restrictive because ν describes the number of filled Landau levels and so can only take integer values. Why should k be quantised in this way?

Rather remarkably, we don't have to assume that k is quantised in this manner; instead, it is *obliged* to take values that are integer multiples of $e^2/\hbar$. This follows from the "almost" part of the almost-gauge invariance of the Chern-Simons term. The quantisation in the Abelian Chern-Simons term (5.6) turns out to be somewhat subtle. (In contrast, it's much more direct to see the corresponding quantisation for the non-Abelian Chern-Simons theories that we introduce in Section 5.4). To see how it arises, it's perhaps simplest to place the theory at finite temperature and compute the corresponding partition function, again with A_μ a source. To explain this, we first need a small aside about how should think about the equilibrium properties of field theories at finite temperature.

5.1.2 An Aside: Periodic Time Makes Things Hot

In this small aside we will look at the connection between the thermal partition function that we work with in statistical mechanics and the quantum partition function that we work with in quantum field theory. To explain this, we're going to go right back to basics. This means the dynamics of a single particle.

Consider a quantum particle of mass m moving in one direction with coordinate q. Suppose it moves in a potential $V(q)$. The statistical mechanics partition function is

$$Z[\beta] = \mathrm{Tr}\, e^{-\beta H} \tag{5.9}$$

where H is, of course, the Hamiltonian operator and $\beta = 1/T$ is the inverse temperature (using conventions with $k_B = 1$). We would like to write down a path integral expression for this thermal partition function.

We're more used to thinking of path integrals for time evolution in quantum mechanics. Suppose the particle sits at some point q_i at time $t = 0$. The Feynman path integral provides an expression for the amplitude for the particle to evolve to position $q = q_f$ at a time t later,

$$\langle q_f | e^{-iHt} | q_i \rangle = \int_{q(0)=q_i}^{q(t)=q_f} \mathcal{D}q\, e^{iS} \tag{5.10}$$

where S is the classical action, given by

$$S = \int_0^t dt' \; \left[\frac{m}{2} \left(\frac{dq}{dt'} \right)^2 - V(q) \right]$$

Comparing (5.9) and (5.10), we see that they look tantalisingly similar. Our task is to use (5.10) to derive an expression for the thermal partition function (5.9). We do this in three steps. We start by getting rid of the factor of i in the quantum mechanics path integral. This is accomplished by Wick rotating, which just means working with the Euclidean time variable

$$\tau = it$$

With this substitution, the action becomes

$$iS = \int_0^{-i\tau} d\tau' \left[-\frac{m}{2}\left(\frac{dq}{d\tau}\right)^2 - V(q)\right] \equiv -S_E$$

where S_E is the *Euclidean action*.

The second step is to introduce the temperature. We do this by requiring the particle propagates for a (Euclidean) time $\tau = \beta$, so that the quantum amplitude becomes,

$$\langle q_f | e^{-H\beta} | q_i \rangle = \int_{q(0)=q_i}^{q(\beta)=q_f} \mathcal{D}q\, e^{-S_E}$$

Now we're almost there. All that's left is to implement the trace. This simply means a sum over a suitable basis of states. For example, if we choose to sum over the initial position, we have

$$\mathrm{Tr}\, \cdot = \int dq_i\ \langle q_i | \cdot | q_i \rangle$$

We see that taking the trace means we should insist that $q_i = q_f$ in the path integral, before integrating over all q_i. We can finally write

$$\begin{aligned} \mathrm{Tr}\, e^{-\beta H} &= \int dq_i\ \langle q_i | e^{-H\beta} | q_i \rangle \\ &= \int dq_i \int_{q(0)=q_i}^{q(\beta)=q_i} \mathcal{D}q\, e^{-S_E} \\ &= \int_{q(0)=q(\beta)} \mathcal{D}q\, e^{-S_E} \end{aligned}$$

The upshot is that we have to integrate over all trajectories with the sole requirement $q(0) = q(\beta)$, with no constraint on what this starting point is. All we have to impose is that the particle comes back to where it started after Euclidean time $\tau = \beta$. This is usually summarised by simply saying that the Euclidean time direction is compact: τ should be thought of as parameterising a circle, with periodicity

$$\tau \equiv \tau + \beta \tag{5.11}$$

Although we've walked through this simple example of a quantum particle, the general lesson that we've seen here holds for all field theories. If you take a quantum field theory that lives on Minkowski space $\mathbf{R}^{d-1,1}$ and want to compute the thermal partition function, then all you have to do is consider the Euclidean path integral, but with the theory now formulated on the Euclidean space $\mathbf{R}^{d-1} \times \mathbf{S}^1$, where the circle is parameterised by $\tau \in [0, \beta)$. There is one extra caveat that you need to know. While all bosonic field are periodic in the time direction (just like $q(\tau)$ in our example above), fermionic fields should be made anti-periodic: they pick up a minus sign as you go around the circle.

All of this applies directly to the thermal partition function for our quantum Hall theory, resulting in an effective action $S_{\text{eff}}[A]$ which itself lives on $\mathbf{R}^2 \times \mathbf{S}^1$. However, there's one small difference for Chern-Simons terms. The presence of the $\epsilon_{\mu\nu\rho}$ symbol in (5.6) means that the action in Euclidean space picks up an extra factor of i. The upshot is that, in both Lorentzian and Euclidean signature, the term in the path integral takes the form $e^{iS_{CS}/\hbar}$. This will be important in what follows.

5.1.3 Quantisation of the Chern-Simons level

We're now in a position to understand the quantisation of the Chern-Simons level k in (5.6). As advertised earlier, we look at the partition function at finite temperature by taking time to be Euclidean $\mathbf{S}^1$, parameterised by τ with periodicity (5.11).

Having a periodic $\mathbf{S}^1$ factor in the geometry allows us to do something novel with gauge transformations, $A_\mu \to A_\mu + \partial_\mu \omega$. Usually, we work with functions $\omega(t, \mathbf{x})$ which are single valued. But that's actually too restrictive: we should ask only that the physical fields are single valued. The electron wavefunction (in the language of quantum mechanics) or field (in the language of, well, fields) transforms as $e^{ie\omega/\hbar}$. So the real requirement is not that ω is single valued, but rather that $e^{ie\omega/\hbar}$ is single valued. And, when the background geometry has a $\mathbf{S}^1$ factor, that allows us to do something novel where the gauge transformations "winds" around the circle, with

$$\omega = \frac{2\pi\hbar\tau}{e\beta} \tag{5.12}$$

which leaves the exponential $e^{ie\omega/\hbar}$ single valued as required. These are sometimes called *large* gauge transformations; the name is supposed to signify that they cannot be continuously connected to the identity. Under such a large gauge transformation, the temporal component of the gauge field is simply shifted by a constant

$$A_0 \to A_0 + \frac{2\pi\hbar}{e\beta} \tag{5.13}$$

Gauge fields that are related by gauge transformations should be considered physically equivalent. This means that we can think of A_0 (strictly speaking, its zero mode) as being a periodic variable, with periodicity $2\pi\hbar/e\beta$, inversely proportional to the radius β of the $\mathbf{S}^1$. Our interest is in how the Chern-Simons term fares under gauge transformations of the type (5.12).

To get something interesting, we'll also need to add one extra ingredient. We think about the spatial directions as forming a sphere $\mathbf{S}^2$, rather than a plane $\mathbf{R}^2$. (This is reminiscent of the kind of set-ups we used in Section 2, where all the general arguments we gave for quantisation involved some change of the background geometry, whether an annulus or torus or lattice). We take advantage of this new geometry by threading a background magnetic flux through the spatial $\mathbf{S}^2$, given by

$$\frac{1}{2\pi}\int_{\mathbf{S}^2} F_{12} = \frac{\hbar}{e} \tag{5.14}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$.This is tantamount to placing a Dirac magnetic monopole inside the $\mathbf{S}^2$. The flux above is the minimum amount allowed by the Dirac quantisation condition. Clearly this experiment is hard to do in practice. It involves building a quantum Hall state on a sphere which sounds tricky. More importantly, it also requires the discovery of a magnetic monopole! However, there should be nothing wrong with doing this in principle. And we will only need the *possibility* of doing this to derive constraints on our quantum Hall system.

We now evaluate the Chern-Simons term (5.6) on a configuration with constant $A_0 = a$ and spatial field strength (5.14). Expanding (5.6), we find

$$S_{CS} = \frac{k}{4\pi}\int d^3x\ A_0F_{12} + A_1F_{20} + A_2F_{01}$$

Now it's tempting to throw away the last two terms when evaluating this on our background. But we should be careful as it's topologically non-trivial configuration. We can safely set all terms with ∂_0 to zero, but integrating by parts on the spatial derivatives we get an extra factor of 2,

$$S_{CS} = \frac{k}{2\pi}\int d^3x\ A_0F_{12}$$

Evaluated on the flux (5.14) and constant $A_0 = a$, this gives

$$S_{CS} = \beta a\frac{\hbar k}{e} \tag{5.15}$$

The above calculation was a little tricky: how do we know that we needed to integrate by parts before evaluating? The reason we got different answers is that we're dealing with a topologically non-trivial gauge field. To do a proper job, we should think about the gauge field as being defined locally on different patches and glued together in an appropriate fashion. (Alternatively, there's a way to think of the Chern-Simons action as living on the boundary of a four dimensional space.) We won't do this proper job here. But the answer (5.15) is the correct one.

Now that we've evaluated the Chern-Simons action on this particular configuration, let's see how it fares under gauge transformations (5.13) which shift A_0. We learn that the Chern-Simons term is not quite gauge invariant after all. Instead, it transforms as

$$S_{CS} \to S_{CS} + \frac{2\pi\hbar^2 k}{e^2}$$

This looks bad. However, all is not lost. Looking back, we see that the Chern-Simons term should really be interpreted as a quantum effective action,

$$Z[A_\mu] = e^{iS_{\rm eff}[A_\mu]/\hbar}$$

It's ok if the Chern-Simons term itself is not gauge invariant, as long as the partition function $e^{iS_{CS}/\hbar}$ is. We see that we're safe provided

$$\frac{\hbar k}{e^2} \in \mathbf{Z}$$

This is exactly the result that we wanted. We now write, $k = e^2\nu/\hbar$ with $\nu \in \mathbf{Z}$. Then the Hall conductivity (5.7) is

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

which is precisely the conductivity seen in the integer quantum Hall effect. Similarly, the charge density (5.8) also agrees with that of the integer quantum Hall effect.

This is a lovely result. We've reproduced the observed quantisation of the integer quantum Hall effect without ever getting our hands dirty. We never needed to discuss what underlying theory we were dealing with. There was no mention of Landau levels, no mention of whether the charge carriers were fermions or bosons, or whether they were free or strongly interacting. Instead, on very general grounds we showed that the Hall conductivity *has* to be quantised. This nicely complements the kinds of microscopic arguments we met in Section 2 for the quantisation of σ_{xy}

Compact vs. Non-Compact

Looking back at the derivation, it seems to rely on two results. The first is the periodic nature of gauge transformations, $e^{ie\omega/\hbar}$, which means that the topologically non-trivial gauge transformations (5.12) are allowed. Because the charge appears in the exponent, an implicit assumption here is that all fields transform with the same charge. We can, in fact, soften this slightly and one can repeat the argument whenever charges are rational multiples of each other. Abelian gauge symmetries with this property are sometimes referred to as *compact*. It is an experimental fact, which we've all known since high school, that the gauge symmetry of Electromagnetism is compact (because the charge of the electron is minus the charge of the proton).

Second, the derivation required there to be a minimum flux quantum (5.14), set by the Dirac quantisation condition. Yet a close inspection of the Dirac condition shows that this too hinges on the compactness of the gauge group. In other words, the compact nature of Electromagnetism is all that's needed to ensure the quantisation of the Hall conductivity.

In contrast, Abelian gauge symmetries which are non-compact — for example, because they have charges which are irrational multiples of each other — cannot have magnetic monopoles, or fluxes of the form (5.14). We sometimes denote their gauge group as $\mathbf{R}$ instead of $U(1)$ to highlight this non-compactness. For such putative non-compact gauge fields, there is no topological restriction on the Hall conductivity.

5.2 The Fractional Quantum Hall Effect

In the last section, we saw very compelling arguments for why the Hall conductivity must be quantised. Yet now that leaves us in a bit of a bind, because we somehow have to explain the fractional quantum Hall effect where this quantisation is not obeyed. Suddenly, the great power and generality of our previous arguments seems quite daunting!

If we want to avoid the conclusion that the Hall conductivity takes integer values, our only hope is to violate one of the assumptions that went into our previous arguments. Yet the only thing we assumed is that there are no dynamical degrees which can affect the low-energy energy physics when the system is perturbed. And, at first glance, this looks rather innocuous: we might naively expect that this is true for any system which has a gap in its spectrum, as long as the energy of the perturbation is smaller than that gap. Moreover, the fractional quantum Hall liquids certainly have a gap. So what are we missing?

What we're missing is a subtle and beautiful piece of physics that has many far reaching consequences. It turns out that there can be degrees of freedom which are gapped, but nonetheless affect the physics at arbitrarily low-energy scales. These degrees of freedom are sometimes called "topological". Our goal in this section is to describe the topological degrees of freedom relevant for the fractional quantum Hall effect.

Let's think about what this means. We want to compute the partition function

$$Z[A_\mu] = \int D(\text{fields})\ e^{iS[\text{fields};A]/\hbar}$$

where A_μ again couples to the fields through the current (5.1). However, this time, we should not integrate out all the fields if we want to be left with a local effective action. Instead, we should retain the topological degrees of freedom. The tricky part is that these topological degrees of freedom can be complicated combinations of the original fields and it's usually very difficult to identify in advance what kind of emergent fields will arise in a given system. So, rather than work from first principles, we will first think about what kinds of topological degrees of freedom may arise. Then we'll figure out the consequences.

In the rest of this section, we describe the low-energy effective theory relevant to Laughlin states with $\nu = 1/m$. In subsequent sections, we'll generalise this to other filling fractions.

5.2.1 A First Look at Chern-Simons Dynamics

In $d = 2+1$ dimensions, the simplest kind of topological field theory involves a $U(1)$ dynamical gauge field a_μ. We stress that this is not the gauge field of electromagnetism, which we'll continue to denote as A_μ. Instead a_μ is an *emergent* gauge field, arising from the collective behaviour of many underlying electrons. You should think of this as something analogous to the way phonons arise as the collective motion of many underlying atoms. We will see the direct relationship between a_μ and the electron degrees of freedom later.

We're used to thinking of gauge fields as describing massless degrees of freedom (at least classically). Indeed, their dynamics is usually described by the Maxwell action,

$$S_{\text{Maxwell}}[a] = -\frac{1}{4g^2}\int d^3x\ f_{\mu\nu}f^{\mu\nu} \tag{5.16}$$

where $f_{\mu\nu} = \partial_\mu a_\nu - \partial_\nu a_\mu$ and g^2 is a coupling constant. The resulting equations of motion are $\partial_\mu f^{\mu\nu} = 0$. They admit wave solutions, pretty much identical to those we met in the *Electromagnetism* course except that in $d = 2+1$ dimensions there is only a single allowed polarisation. In other words, $U(1)$ Maxwell theory in $d = 2+1$ dimension describes a single massless degree of freedom.

However, as we've already seen, there is a new kind of action that we can write down for gauge fields in $d = 2+1$ dimensions. This is the Chern-Simons action

$$S_{CS}[a] = \frac{k}{4\pi} \int d^3x \; \epsilon^{\mu\nu\rho} a_\mu \partial_\nu a_\rho \tag{5.17}$$

The arguments of the previous section mean that k must be integer (in units of $e^2/\hbar$) if the emergent $U(1)$ symmetry is compact.

Let's see how the Chern-Simons term changes the classical and quantum dynamics[46]. Suppose that we take as our action the sum of the two terms

$$S = S_{\text{Maxwell}} + S_{CS}$$

The equation of motion for a_μ now becomes

$$\partial_\mu f^{\mu\nu} + \frac{kg^2}{4\pi} \epsilon^{\nu\rho\sigma} f_{\rho\sigma} = 0$$

Now this no longer describes a massless photon. Instead, any excitation decays exponentially. Solving the equations is not hard and one finds that the presence of the Chern-Simons term gives the photon mass M. Equivalently, the spectrum has an energy gap $E_{\text{gap}} = Mc^2$. A short calculation shows that it is given by

$$E_{\text{gap}} = \frac{kg^2}{2\pi}$$

(Note: you need to divide by $\hbar$ on the right-hand side to get something of the right dimension).

In the limit $g^2 \to \infty$, the photon becomes infinitely massive and we're left with no physical excitations at all. This is the situation described by the Chern-Simons theory (5.17) alone. One might wonder what the Chern-Simons theory can possibly describe given that there are no propagating degrees of freedom. The purpose of this section is to answer this!

Chern-Simons Terms are Topological

Before we go on, let us point out one further interesting and important property of (5.17): it doesn't depend on the metric of the background spacetime manifold. It depends only on the topology of the manifold. To see this, let's first look at the

[46] An introduction to Chern-Simons theory can be found in G. Dunne, "*Aspects of Chern-Simons Theory*", hep-th/9902115.

Maxwell action (5.16). If we are to couple this to a background metric $g_{\mu\nu}$, the action becomes

$$S_{\text{Maxwell}} = -\frac{1}{4g^2}\int d^3x\ \sqrt{-g}\, g^{\mu\rho}g^{\nu\sigma}f_{\mu\nu}f_{\rho\sigma}$$

We see that the metric plays two roles: first, it is needed to raise the indices when contracting $f_{\mu\nu}f^{\mu\nu}$; second it provides a measure $\sqrt{-g}$ (the volume form) which allows us to integrate in a diffeomorphism invariant way.

In contrast, neither of these are required when generalising (5.17) to curved spacetime. This is best stated in the language of differential geometry: $a \wedge da$ is a 3-form, and we can quite happily integrate this over any three-dimensional manifold

$$S_{CS} = \frac{k}{4\pi}\int a \wedge da$$

The action is manifestly independent of the metric. In particular, recall from our *Quantum Field Theory* lectures, that we can compute the stress-energy tensor of any theory by differentiating with respect to the metric,

$$T^{\mu\nu} = \frac{2}{\sqrt{-g}}\frac{\partial \mathcal{L}}{\partial g_{\mu\nu}}$$

For Chern-Simons theory, the stress-energy tensor vanishes. This means that the Hamiltonian vanishes. It is an unusual kind of theory.

However, will see in Section 5.2.3 that the topology of the underlying manifold does play an important role in Chern-Simons theory. This will be related to the ideas of topological order that we introduced in Section 3.2.5. Ultimately, it is this topological nature of the Chern-Simons interaction which means that we can't neglect it in low-energy effective actions.

5.2.2 The Effective Theory for the Laughlin States

Now we're in a position to describe the effective theory for the $\nu = 1/m$ Laughlin states. These Hall states have a single emergent, compact $U(1)$ gauge field a_μ. This is a dynamical field, but we should keep it in our effective action. The partition function can then be written as

$$Z[A_\mu] = \int \mathcal{D}a_\mu\ e^{iS_{\text{eff}}[a;A]/\hbar}$$

where $\mathcal{D}a_\mu$ is short-hand for all the usual issues involving gauge-fixing that go into defining a path integral for a gauge field.

Our goal now is to write down $S_{\text{eff}}[a; A]$. However, to get something interesting we're going to need a coupling between A_μ and a_μ. Yet we know that A_μ has to couple to the electron current J^μ. So if this is going to work at all, we're going to have to find a relationship between a_μ and J^μ.

Thankfully, conserved currents are hard to come by and there's essentially only one thing that we can write down. The current is given by

$$J^\mu = \frac{e^2}{2\pi\hbar}\epsilon^{\mu\nu\rho}\,\partial_\nu a_\rho \tag{5.18}$$

The conservation of the current, $\partial_\mu J^\mu = 0$, is simply an identity when written like this. This relation means that the magnetic flux of a_μ is interpreted as the electric charge that couples to A_μ. The normalisation follows directly if we take the emergent $U(1)$ gauge symmetry to be compact, coupling to particles with charge e. In this case, the minimum allowed flux is given by the Dirac quantisation condition

$$\frac{1}{2\pi}\int_{\mathbf{S}^2} f_{12} = \frac{\hbar}{e} \tag{5.19}$$

The relationship (5.18) then ensures that the minimum charge is $\int J^0 = e$ as it should be. (Picking different signs of the flux f_{12} corresponds to electrons and holes in the system).

We then postulate the following effective action,

$$S_{\text{eff}}[a; A] = \frac{e^2}{\hbar}\int d^3x\;\; \frac{1}{2\pi}\,\epsilon^{\mu\nu\rho}A_\mu\partial_\nu a_\rho - \frac{m}{4\pi}\,\epsilon^{\mu\nu\rho}a_\mu\partial_\nu a_\rho + \ldots \tag{5.20}$$

The first term is a "mixed" Chern-Simons term which comes from the $A_\mu J^\mu$ coupling; the second term is the simplest new term that we can write down. By the same arguments that we used before, the level must be integer: $m \in \mathbf{Z}$. As we will see shortly, it is no coincidence that we've called this integer m. The $\ldots$ above include more irrelevant terms, including the Maxwell term (5.16). At large distances, none of them will play any role and we will ignore them in what follows. We could also add a Chern-Simons $\epsilon^{\mu\nu\rho}A_\mu\partial_\nu A_\rho$ for A itself but we've already seen what this does: it simply gives an integer contribution to the Hall conductivity. Setting the coefficient of this term to zero will be equivalent to working in the lowest Landau level.

Let's start by computing the Hall conductivity. The obvious way to do this is to reduce the effective action to something which involves only A by explicitly integrating out the dynamical field a. Because the action is quadratic in a, this looks as if it's going to be easy to do. We would naively just replace such a field with its equation of motion which, in this case, is

$$f_{\mu\nu} = \frac{1}{m} F_{\mu\nu} \tag{5.21}$$

The solution to this equation is $a_\mu = A_\mu/m$ (up to a gauge transformation). Substituting this back into the action (5.20) gives

$$S_{\text{eff}}[A] = \frac{e^2}{2\pi} \int d^3x \;\; \frac{1}{4\pi m} \, \epsilon^{\mu\nu\rho} A_\mu \partial_\nu A_\rho \tag{5.22}$$

This is now the same kind of action (5.6) that we worked with before and we can immediately see that the Hall conductivity is

$$\sigma_{xy} = \frac{e^2}{2\pi\hbar} \frac{1}{m} \tag{5.23}$$

as expected for the Laughlin state.

Although we got the right answer for the Hall conductivity, there's something very fishy about our derivation. The kind of action (5.22) that we ended up lies in the class that we previously argued wasn't allowed by gauge invariance if our theory is defined on a sphere! Our mistake was that we were too quick in the integrating out procedure. The gauge field a_μ is constrained by the Dirac quantisation condition (5.19). But this is clearly incompatible with the equation of motion (5.21) whenever F also has a single unit of flux (5.14). In fact, it had to be this way. If it was possible to integrate out a_μ, then it couldn't have been playing any role in the first place!

Nonetheless, the final answer (5.23) for the Hall conductivity is correct. To see this, just consider the theory on the plane with $F_{12} = 0$ where there are no subtleties with (5.21) and the calculation above goes through without a hitch. However, whenever we want to compute something where monopoles are important, we can't integrate out a_μ. Instead, we're obliged to work with the full action (5.20).

Quasi-Holes and Quasi-Particles

The action (5.20) describes the quantum Hall state at filling $\nu = 1/m$. Let's now add something new to this. We will couple the emergent gauge field a_μ to its own current, which we call j^μ, through the additional term

$$\Delta S = \int d^3x \; a_\mu j^\mu$$

To ensure gauge invariance, j^μ must be conserved: $\partial_\mu j^\mu = 0$. We will now show that the current j^μ describes the quasi-holes and quasi-particles in the system.

First, we'll set the background gauge field A_μ to zero. (It is, after all, a background parameter at our disposal in this framework). The equation of motion for a_μ is then

$$\frac{e^2}{2\pi\hbar} f_{\mu\nu} = \frac{1}{m}\epsilon_{\mu\nu\rho} j^\rho \tag{5.24}$$

The simplest kind of current we can look at is a static charge which we place at the origin. This is described by $j^1 = j^2 = 0$ and $j^0 = e\delta^2(x)$. Note that the fact these particles have charge e under the gauge field a_μ is related to our choice of Dirac quantisation (5.19). The equation of motion above then becomes

$$\frac{1}{2\pi} f_{12} = \frac{\hbar}{em}\,\delta^2(x) \tag{5.25}$$

This is an important equation. We see that the effect of the Chern-Simons term is to attach flux $\hbar/em$ to each particle of charge e. From this we'll see that the particle has both the fractional charge and fractional statistics appropriate for the Laughlin state. The fractional charge follows immediately by looking at the electron current J^μ in (5.18) which, in this background, is

$$J^0 = \frac{e^2}{2\pi\hbar} f_{12} = \frac{e}{m}\delta^2(x)$$

This, of course, is the current appropriate for a stationary particle of electric charge e/m.

Note: the flux attachment (5.25) doesn't seem compatible with the Dirac quantisation condition (5.19). Indeed, if we were on a spatial sphere $\mathbf{S}^2$ we would be obliged to add m quasi-particles, each of charge e/m. However, these particles can still roam around the sphere independently of each other so they should still be considered as individual object. On the plane $\mathbf{R}^2$, we need not be so fussy: if we don't have a multiple of m quasi-holes, we can always think of the others as being somewhere off at infinity.

To see how the fractional statistics emerge, we just need the basic Aharonov-Bohm physics that we reviewed in Section 1.5.3. Recall that a particle of charge q moving around a flux Φ picks up a phase $e^{iq\Phi/\hbar}$. But because of flux attachment (5.25), our quasi-particles necessarily carry both charge $q = e$ and flux $\Phi = 2\pi\hbar/em$. If we move one particle all the way around another, we will get a phase $e^{iq\Phi/\hbar}$. But the statistical phase is defined by exchanging particles, which consists of only half an orbit (followed by a translation which contributes no phase). So the expected statistical phase is $e^{i\pi\alpha} = e^{iq\Phi/2\hbar}$. For our quasi-holes, with $q = e$ and $\Phi = 2\pi\hbar/em$, we get

$$\alpha = \frac{1}{m}$$

which is indeed the expected statistics of quasi-holes in the Laughlin state.

The attachment of the flux to the quasi-hole is reminiscent of the composite fermion ideas that we met in Section 3.3.2, in which we attached vortices (which were zeros of the wavefunction) to quasi-holes.

Fractional Statistics Done Better

The above calculation is nice and quick and gives the right result. But there's a famously annoying factor of 2 that we've swept under the rug. Here's the issue. As the charge q in the first particle moved around the flux Φ in the second, we picked up a phase $e^{iq\Phi/\hbar}$. But you might think that the flux Φ of the first particle also moved around the charge q of the second. So surely this should give another factor of $e^{iq\Phi/\hbar}$. Right? Well, no. To see why, it's best to just do the calculation.

For generality, let's take N particles sitting at positions $\mathbf{x}_a(t)$ which, as the notation shows, we allow to change with time. The charge density and currents are

$$j^0(\mathbf{x},t) = e\sum_{a=1}^{N}\delta^2(\mathbf{x}-\mathbf{x}_a(t)) \quad \text{and} \quad \mathbf{j}(\mathbf{x},t) = e\sum_{a=1}^{N}\dot{\mathbf{x}}_a\,\delta^2(\mathbf{x}-\mathbf{x}_a(t))$$

The equation of motion (5.24) can be easily solved even in this general case. If we work in the Coulomb gauge $a_0 = 0$ with $\partial_i a_i = 0$ (summing over spatial indices only), the solution is given by

$$a_i(\mathbf{x},t) = \frac{\hbar}{em}\sum_{a=1}^{N}\epsilon^{ij}\,\frac{x^j - x_a^j(t)}{|\mathbf{x}-\mathbf{x}_a(t)|^2} \tag{5.26}$$

This follows from the standard methods that we know from our *Electromagnetism* course, but this time using the Green's function for the Laplacian in two dimensions: $\nabla^2\log|\mathbf{x}-\mathbf{y}| = 2\pi\delta^2(\mathbf{x}-\mathbf{y})$. This solution is again the statement that each particle carries flux $\hbar/em$. However, we can also use this solution directly to compute the phase change when one particle – say, the first one – is transported along a curve C. It is simply

$$\exp\left(ie\oint_C \mathbf{a}\cdot d\mathbf{x}_1\right)$$

If the curve C encloses one other particle, the resulting phase change can be computed to be $e^{2\pi i/m}$. As before, if we exchange two particles, we get half this phase, or $e^{\imath\pi\alpha} = e^{i\pi/m}$. This, of course, is the same result we got above.

It's worth pointing out that this Chern-Simons computation ended up looking exactly the same as the original Berry phase calculation for the Laughlin wavefunctions that we saw in Section 3.2.3. For example, the connection (5.26) is identical to the relevant part of the Berry connections (3.25) and (3.26). (The seeming difference in the factor of 2 can be traced to our previous normalisation for complex connections).

Breathing Life into the Quasi-Holes

In the calculations above, we've taken j^μ to be some fixed, background current describing the quasi-particles. But the framework of effective field theory also allows us to make the quasi-particles dynamical. We simply need to introduce a new bosonic field ϕ and include it in the effective action, coupled minimally to a_μ. We then endow ϕ with its own dynamics. Exactly what dynamics we choose is up to us at this point. For example, if we wanted the quasi-holes to have a relativistic dispersion relation, we would introduce the action

$$S_{\text{eff}}[a,\phi] = \int d^3x\ \frac{e^2 m}{4\pi\hbar}\epsilon^{\mu\nu\rho}a_\mu\partial_\nu a_\rho + |\mathcal{D}_\mu\phi|^2 - V(\phi)$$

where the relativistic form of the action also implies that ϕ will describe both particle and anti-particle (i.e. hole) excitations. Here $V(\phi)$ is a potential that governs the mass and self-interactions of interactions of the quasi-particles. Most important, the covariant derivative $\mathcal{D}_\mu = \partial_\mu - iea_\mu$ includes the coupling to the Chern-Simons field. By the calculations above, this ensures that the excitations of ϕ will have the correct anyonic statistics to describe quasi-particles, even though the field ϕ itself is bosonic.

We'll see a different way to make the current j^μ dynamical in Section 5.2.4 when we discuss other filling fractions.

5.2.3 Chern-Simons Theory on a Torus

In Section 3.2.5, we argued that if we place a fractional quantum Hall state on a compact manifold, then the number of ground states depends on the topology of that manifold. In particular, we showed that the existence of anyons alone was enough to ensure m ground states on a torus and m^g ground states on a genus-g surface. This is the essence of what's known as *topological order*.

Here we show how this is reproduced by the Chern-Simons theory. If we live on the plane $\mathbf{R}^2$ or the sphere $\mathbf{S}^2$, then Chern-Simons theory has just a single state. But if we change the background manifold to be anything more complicated, like a torus, then there is a degeneracy of ground states.

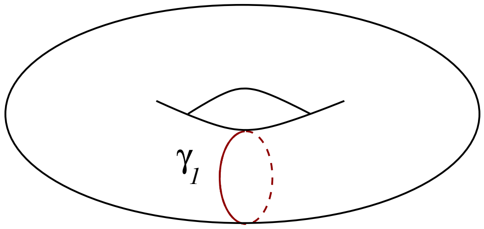


Figure 42:

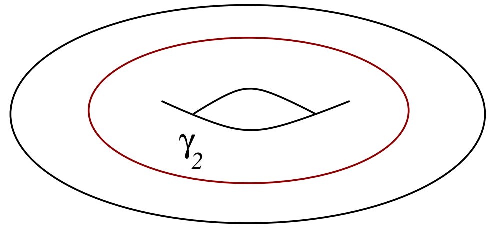


Figure 43:

To see this effect, we can turn off the background sources and focus only on the dynamical part of the effective theory,

$$S_{CS} = \frac{e^2}{\hbar} \int d^3x \; \frac{m}{4\pi} \epsilon^{\mu\nu\rho} \, a_\mu \partial_\nu a_\rho \tag{5.27}$$

The equation of motion for a_0, known, in analogy with electromagnetism, as Gauss' law, is

$$f_{12} = 0$$

Although this equation is very simple, it can still have interesting solutions if the background has some non-trivial topology. These are called, for obvious reason, *flat connections*. It's simple to see that such solutions exist on the torus $\mathbf{T}^2$, where one example is to simply set each a_i to be constant. Our first task is to find a gauge-invariant way to parameterise this space of solutions.

We'll denote the radii of the two circles of the torus $\mathbf{T}^2 = \mathbf{S}^1 \times \mathbf{S}^1$ as R_1 and R_2. We'll denote two corresponding non-contractible curves shown in the figure as γ_1 and γ_2. The simplest way to build a gauge invariant object from a gauge connection is to integrate

$$w_i = \oint_{\gamma_i} dx^j \, a_j$$

This is invariant under most gauge transformations, but not those that wind around the circle. By the same kind of arguments that led us to (5.13), we can always construct gauge transformations which shift $a_j \rightarrow a_j + \hbar/eR_j$, and hence $w_i \rightarrow w_i + 2\pi\hbar/e$. The correct gauge invariant objects to parameterise the solutions are therefore the Wilson loops

$$W_i = \exp\left(i\frac{e}{\hbar} \oint_{\gamma_i} a_j dx^j \right) = e^{iew_i/\hbar}$$

Because the Chern-Simons theory is first order in time derivatives, these Wilson loops are really parameterising the phase space of solutions, rather than the configuration space. Moreover, because the Wilson lines are complex numbers of unit modulus, the phase space is compact. On general grounds, we would expect that when we quantise a compact phase space, we get a finite-dimensional Hilbert space. Our next task is to understand how to do this.

The canonical commutation relations can be read off from the Chern-Simons action (5.27)

$$[a_1(x), a_2(x')] = \frac{2\pi i}{m}\frac{\hbar^2}{e^2}\,\delta^2(x-x') \quad \Rightarrow \quad [w_1, w_2] = \frac{2\pi i}{m}\frac{\hbar^2}{e^2}$$

The algebraic relation obeyed by the Wilson loops then follows from the usual Baker-Campbell-Hausdorff formula,

$$e^{iew_1/\hbar}e^{iew_2/\hbar} = e^{e^2[w_1,w_2]/2\hbar^2}\,e^{ie(w_1+w_2)/\hbar}$$

Or, in other words,

$$W_1W_2 = e^{2\pi i/m}\,W_2W_1 \tag{5.28}$$

But this is exactly the same as the algebra (3.33) that we met when considering anyons on a torus! This is not surprising: one interpretation of the Wilson loop is for a particle charged under e to propagate around the cycle of the torus. And that's exactly how we introduced the operators T_i that appear in (3.33).

From Section 3.2.5, we know that the smallest representation of the algebra (5.28) has dimension m. This is the number of ground states of the Chern-Simons theory on a torus. The generalisation of the above calculation to a genus-g surface gives a ground state degeneracy of m^g.

5.2.4 Other Filling Fractions and K-Matrices

It's straightforward to generalise the effective field theory approach to other filling fractions. We'll start by seeing how the hierarchy of states naturally emerges. To simplify the equations in what follows, we're going to use units in which $e = \hbar = 1$. (Nearly all other texts resort to such units long before now!)

The Hierarchy

The effective field theory for the Laughlin states that we saw above can be summarised as follows: we write the electron current as

$$J^\mu = \frac{1}{2\pi}\epsilon^{\mu\nu\rho}\,\partial_\nu a_\rho \tag{5.29}$$

where a_μ is an emergent field. We then endow a_μ with a Chern-Simons term.

Now we'd like to repeat this to implement the hierarchy construction described in Section 3.3.1 in which the quasi-particles themselves form a new quantum Hall state. But that's very straightforward. We simply write the quasi-particle current j^μ as

$$j^\mu = \frac{1}{2\pi}\epsilon^{\mu\nu\rho}\,\partial_\nu \tilde{a}_\rho \tag{5.30}$$

where $\tilde{a}_\mu$ is a second emergent gauge field whose dynamics are governed by a second Chern-Simons term. The final action is

$$S_{\text{eff}}[a,\tilde{a};A] = \int d^3x\ \ \frac{1}{2\pi}\,\epsilon^{\mu\nu\rho}A_\mu\partial_\nu a_\rho - \frac{m}{4\pi}\,\epsilon^{\mu\nu\rho}a_\mu\partial_\nu a_\rho + \frac{1}{2\pi}\epsilon^{\mu\nu\rho}a_\mu\partial_\nu\tilde{a}_\rho - \frac{\tilde{m}}{4\pi}\,\epsilon^{\mu\nu\rho}\tilde{a}_\mu\partial_\nu\tilde{a}_\rho$$

To compute the Hall conductivity, we can first integrate out $\tilde{a}$ and then integrate out a. We find that this theory describes a Hall state with filling fraction

$$\nu = \frac{1}{m - \frac{1}{\tilde{m}}}$$

When $\tilde{m}$ is an even integer, this coincides with our expectation (3.34) for the first level of the hierarchy.

We can also use this approach to quickly compute the charge and statistics of quasi-particles in this state. There are two such quasi-holes, whose currents couple to a and $\tilde{a}$ respectively. For a static quasi-hole which couples to a, the equations of motion read

$$mf_{12} - \tilde{f}_{12} = 2\pi\delta^2(x) \quad \text{and} \quad \tilde{m}\tilde{f}_{12} - f_{12} = 0 \quad \Rightarrow \quad f_{12} = \frac{2\pi}{m - 1/\tilde{m}}\delta^2(x)$$

while, if the quasi-hole couples to $\tilde{a}$, the equations of motion are

$$mf_{12} - \tilde{f}_{12} = 0 \quad \text{and} \quad \tilde{m}\tilde{f}_{12} - f_{12} = 2\pi\delta^2(x) \quad \Rightarrow \quad f_{12} = \frac{2\pi}{m\tilde{m} - 1}\delta^2(x)$$

The coefficients of the right-hand side of the final equations tell us the electric charge.

For example, the $\nu = 2/5$ state has $m = 3$ and $\tilde{m} = 2$. The resulting charges of the quasi-holes are $e^\star = 2/5$ and $e^\star = 1/5$. This has been confirmed experimentally. Using the results from either Section 3.2.5 or Section 5.2.3, we learn that the the $\nu = 2/5$ state has a 5-fold degeneracy on the torus.

Now it's obvious how to proceed: the quasi-particles of the new state are described by a current $j_2(x)$ which couples to $\tilde{a}_\mu$. We write this in the form (5.30) and introduce the new, third, emergent gauge field with a Chern-Simons term. And so on and so on. The resulting states have filling fraction

$$\nu = \cfrac{1}{m \pm \cfrac{1}{\tilde{m}_1 \pm \cfrac{1}{\tilde{m}_2 \pm \cdots}}}$$

which is the result that we previously stated (3.35) without proof.

K-Matrices

Using these ideas, we can now write down the effective theory for the most general Abelian quantum Hall state. We introduce N emergent gauge fields a^i_μ, with $i = 1, \ldots, N$. The most general theory is

$$S_K[a^i, A] = \int d^3x\; \frac{1}{4\pi} K_{ij} \epsilon^{\mu\nu\rho} a^i_\mu \partial_\nu a^j_\rho + \frac{1}{2\pi} t_i \epsilon^{\mu\nu\rho} A_\mu \partial_\nu a^i_\nu \tag{5.31}$$

It depends on the K*-matrix*, K_{ij}, which specifies the various Chern-Simons couplings, and the charge vector t_i which specifies the linear combination of currents that is to be viewed as the electron current. We could also couple different quasi-holes currents to other linear combinations of the a^i

The K-matrix and t-vector encode much of the physical information that we care about. The Hall conductance is computed by integrating out the gauge fields and is given by

$$\sigma_{xy} = (K^{-1})^{ij} t_i t_j$$

the charge of the quasi-hole which couples to the gauge field a^i is

$$(e^\star)^i = (K^{-1})^{ij} t_j$$

and the statistics between quasi-holes that couple to a^i and those that couple to a^j is

$$\alpha^{ij} = (K^{-1})^{ij}$$

One can also show, by repeating the kinds of arguments we gave in Section 5.2.3, that the ground state degeneracy on a genus-g surface is $|\mathrm{det} K|^g$.

We've already met the K-matrix associated to the hierarchy of states. It is

$$K = \begin{pmatrix} m & -1 & 0 & \dots \\ -1 & \tilde{m}_1 & -1 & \\ 0 & -1 & \tilde{m}_2 & \\ \vdots & & & \ddots \end{pmatrix} \quad \text{and} \quad t = (1, 0, 0 \dots)$$

But we can also use the K-matrix approach to describe other Hall states. For example, the (m_1, m_2, n) Halperin states that we met in Section 3.3.4 have K-matrices given by

$$K = \begin{pmatrix} m_1 & n \\ n & m_2 \end{pmatrix} \quad \text{and} \quad t = (1, 1)$$

Using our formula above, we find that the filling fraction is

$$\nu = (K^{-1})^{ij} t_i t_j = \frac{m_1 + m_2 - 2n}{m_1 m_2 - n^2}$$

in agreement with our earlier result (3.46). The ground state degeneracy on a torus is $|m_1 m_2 - n^2|$.

Restricting now to the (m, m, n) states, we can compute the charges and statistics of the two quasi-holes. From the formulae above, we can read off straightaway that the two quasi-holes have charges $e^\star = 1/(m+n)$ and $\alpha = m/(m^2 - n^2)$. We can also take appropriate bound states of these quasi-holes that couple to other linear combinations of a^1 and a^2

Relating Different K-Matrices

Not all theories (5.31) with different K-matrices and t-vectors describe different physics. We could always rewrite the theory in terms of different linear combinations of the gauge fields. After this change of basis,

$$K \to SKS^T \quad \text{and} \quad t \to St \tag{5.32}$$

However, there's an extra subtlety. The gauge fields in (5.31) are all defined such that their fluxes on a sphere are integer valued: $\frac{1}{2\pi}\int_{\mathbf{S}^2} f^i_{12} \in \mathbf{Z}$, just as in (5.19). This should be maintained under the change of basis. This holds as long as the matrix S above lies in $SL(N, \mathbf{Z})$.

The pair (K, t), subject to the equivalence (5.32), are almost enough to classify the possible Abelian quantum states. It turns out, however, that there's one thing missing. This is known as the *shift*. It is related to the degeneracy when the Hall fluid is placed on manifolds of different topology; you can read about this in the reviews by Wen or Zee. More recently, it's been realised that the shift is also related to the so-called Hall viscosity of the fluid.

5.3 Particle-Vortex Duality

The effective field theories that we've described above were not the first attempt to use Chern-Simons theory as a description of the quantum Hall effect. Instead, the original attempts tried to write down local order parameters for the quantum Hall states and build a low-energy effective theory modelled on the usual Ginzburg-Landau approach that we met in the *Statistical Physics* lectures.

It's now appreciated that the more subtle topological aspects of the quantum Hall states that we've described above are not captured by a Ginzburg-Landau theory. Nonetheless, this approach provides a framework in which many detailed properties of the quantum Hall states can be computed. We won't provide all these details here and this section will be less comprehensive than others. Its main purpose is to explain how to construct these alternative theories and provide some pointers to the literature. Moreover, we also take this opportunity to advertise a beautiful property of quantum field theories in $d = 2 + 1$ dimensions known as particle-vortex duality.

5.3.1 The XY-Model and the Abelian-Higgs Model

In $d = 2 + 1$ dimensional field theories, there are two kinds of particle excitations that can appear. The first kind is the familiar excitation that we get when we quantise a local field. This is that kind that we learned about in our *Quantum Field Theory* course. The second kind of particle is a *vortex*, defined by the winding of some local order parameter. These arise as solitons of the theory.

Often in $d = 2 + 1$ dimensions, it's possible to write down two very different-looking theories which describe the same physics. This is possible because the particles of one theory are related to the vortices of the other, and vice versa. We start by explaining how this works in the simplest example, first proposed in the 70's by Peskin and early '80's by Dasgupta and Halperin.

Theory A: The XY-Model

Our first theory consists only of a complex scalar field ϕ with action

$$S_A = \int d^3x \; |\partial_\mu \phi|^2 - a|\phi|^2 - b|\phi|^4 + \ldots \tag{5.33}$$

The theory has a global $U(1)$ symmetry which acts by rotations of the form $\phi \to e^{i\theta}\phi$. The different phases of this theory, and the corresponding physical excitations, can be characterised by symmetry breaking of this $U(1)$. There are three different possibilities which we'll characterise by the sign of a (assuming that $b > 0$),

- $a > 0$: In this phase, the $U(1)$ symmetry is unbroken and the ϕ excitations are massive.

- $a < 0$: In this phase, ϕ gets a vacuum expectation value and the $U(1)$ global symmetry is broken. We can write $\phi = \rho e^{i\sigma}$. The fluctuations of ρ are massive, while the σ field is massless: it is the Goldstone mode for the broken $U(1)$. This phase is sometimes called the "*XY model*" (as it also arises from lattice models of spins which can rotate freely in the (x, y)-plane).

  In this phase, the theory also has vortex excitations. These arise from the phase of ϕ winding asymptotically. The winding is measured by

  $$\oint dx^i \; \partial_i \sigma = 2\pi n$$

  with $n \in \mathbf{Z}$ countrs the number of vortices (or anti-vortices for $n < 0$). Note that n is quantised for topological reasons. These vortices are gapped. Indeed, if you compute their energy from the action (5.33), you'll find that it is logarithmically divergent. Said another way, there is a logarithmically increasing attractive force between a vortex and an anti-vortex. The vortices are sometimes said to be "logarithmically confined".

- $a = 0$: Lying between the two phases above is a critical point. We are being a little careless in describing this as $a = 0$; strictly, you should tune both a and the other parameters to sit at this point. Here, the low-energy dynamics is described by a conformal field theory.

We now compare this to the physics that arises in a very different theory:

Theory B: The Abelian-Higgs Model

Our second theory again consists of a complex scalar field, which we now call $\tilde{\phi}$. This time the scalar is coupled to a dynamical gauge field α_μ. The action is

$$S_B = \int d^3x \; -\frac{1}{4g^2}\tilde{f}_{\mu\nu}\tilde{f}^{\mu\nu} + |\mathcal{D}_\mu\tilde{\phi}|^2 - a'|\tilde{\phi}|^2 - b'|\tilde{\phi}|^4 + \ldots \tag{5.34}$$

with $\tilde{f}_{\mu\nu} = \partial_\mu\alpha_\nu - \partial_\nu\alpha_\mu$. At first glance, Theory A and Theory B look very different. Nonetheless, as we now explain, they describe the same physics. Let's start by matching the symmetries.

Theory B clearly has a $U(1)$ gauge symmetry. This has no counterpart in Theory A but that's ok because gauge symmetries aren't real symmetries: they are merely redundancies in our description of the system. It's more important to match the global symmetries. We've seen that Theory A has a $U(1)$ global symmetry. But there is also a less obvious global symmetry in Theory B, with the current given by

$$j^\mu = \frac{1}{2\pi}\epsilon^{\mu\nu\rho}\partial_\nu\alpha_\rho \tag{5.35}$$

This is the kind of current that we were playing with in our theories of the quantum Hall effect. The conserved charge is the magnetic flux associated to the $U(1)$ gauge symmetry. This is to be identified with the global $U(1)$ symmetry in Theory A.

Now let's look at the different phases exhibited by Theory B. Again, assuming that $b' > 0$, there are three phases depending on the sign of a',

- $a' > 0$: In this phase, the $\tilde{\phi}$ fields are massive and the $U(1)$ gauge symmetry is unbroken. Correspondingly, there is a massless photon in the spectrum. This is usually referred to as the *Coulomb phase*. However, in $d = 2+1$ dimensions, the photon carries only a single polarisation state and can be alternatively described by a scalar field, usually referred to as the *dual photon*, σ. We can implement the change of variables in the path integral if we ignore the coupling to the $\tilde{\phi}$ fields. We can then replace the integration over α_μ with an integration over the field strength $\tilde{f}_{\mu\nu}$ then, schematically (ignoring issues of gauge fixing) the partition function reads

$$\begin{aligned} Z &= \int \mathcal{D}\alpha \; \exp\left(i\int d^3x \; -\frac{1}{4g^2}\tilde{f}_{\mu\nu}\tilde{f}^{\mu\nu}\right) \\ &= \int \mathcal{D}\tilde{f}\mathcal{D}\sigma \; \exp\left(i\int d^3x \; -\frac{1}{4g^2}\tilde{f}_{\mu\nu}\tilde{f}^{\mu\nu} + \frac{1}{2\pi}\sigma\epsilon^{\mu\nu\rho}\partial_\mu\tilde{f}_{\nu\rho}\right) \end{aligned}$$

Here σ is playing the role of a Lagrange multiplier whose role is to impose the Bianchi identity $\epsilon^{\mu\nu\rho}\partial_\mu \tilde{f}_{\nu\rho} = 0$. If the field strength obeys the Dirac quantisation condition, then σ has periodicity 2π. Now we integrate out the field strength, leaving ourselves only with an effective action for σ,

$$Z = \exp\left(i\int d^3x \ \frac{g^2}{2\pi}\partial_\mu\sigma\partial^\mu\sigma\right)$$

This is the dual photon. It is related to the original field strength by the equation of motion

$$\tilde{f}^{\mu\nu} = \frac{g^2}{\pi}\epsilon^{\mu\nu\rho}\partial_\rho\sigma$$

Note that the current (5.35) can be easily written in terms of the dual photon: it is

$$j_\mu = \frac{g^2}{\pi}\partial_\mu\sigma$$

Another way of saying this is that the global $U(1)$ symmetry acts by shifting the value of the dual photon: $\sigma \to \sigma + \text{const}$.

The upshot of this is that the global $U(1)$ symmetry is spontaneously broken in this phase. This means that we should identify the Coulomb phase of Theory B with the $a < 0$ phase of Theory A. The dual photon σ can be viewed as the Goldstone mode of this broken symmetry. This is to be identified with the Goldstone mode of the $a < 0$ phase of Theory A. (Indeed, we even took the liberty of giving the two Goldstone modes the same name.)

The charged $\tilde{\phi}$ fields are massive in this phase. These are to be identified with the vortices of the $a < 0$ phase of Theory A. As a check, note that the $\tilde{\phi}$ excitations interact through the Coulomb force which, in $d = 2 + 1$ dimensions, results in a logarithmically confining force between charges of opposite sign, just like the vortices of Theory A.

- $a' < 0$: In this phase $\tilde{\phi}$ gets an expectation value and the $U(1)$ gauge symmetry is broken. Now the photon gets a mass by the Higgs mechanism and all excitations are gapped. This is the *Higgs* phase of the theory.

  The global $U(1)$ symmetry is unbroken in this phase. This means that we should identify the Higgs phase of Theory B with the gapped $a > 0$ phase of Theory A.

The breaking of the $U(1)$ gauge symmetry means that there are vortex solutions in the Higgs phase. These are defined by the asymptotic winding of the expectation value of $\tilde{\phi}$. The resulting solutions exhibit some nice properties[47]. First, unlike the global vortices of Theory A, vortices associated to a gauge symmetry have finite mass. Second, they also carry quantised magnetic flux

$$\oint dx^i \partial_i \arg(\tilde{\phi}) = \frac{1}{2\pi} \int d^2x \; \tilde{f}_{12} = n'$$

where $n' \in \mathbf{Z}$ is the number of vortices. The fact that these vortices carry magnetic flux means that they are charged under the current (5.35). These vortices are identified with the ϕ excitations of Theory A in the $a > 0$ phase.

- $a' = 0$: Lying between these two phases, there is again a quantum critical point. Numerical simulations show that this is the same quantum critical point that exists in Theory A.

We can see that, viewed through a blurred lens, the theories share the same phase diagram. Roughly, the parameters of are related by

$$a \approx -a'$$

Note, however, that we're only described how qualitative features match. If you want to go beyond this, and see how the interactions match in detail then it's much harder and you have to worry about all the ... interactions in the two theories that we didn't write down. (For what it's worth, you can go much further in supersymmetric theories where the analog of this duality is referred to as *mirror symmetry*).

The qualitative level of the discussion above will be more than adequate for our purposes. Our goal now is to apply these ideas to the effective field theories that we previously wrote down for the fractional quantum Hall effect.

5.3.2 Duality and the Chern-Simons Ginzburg-Landau Theory

So far, the duality that we've described has nothing to do with the quantum Hall effect. However, it's simple to tinker with this duality to get the kind of theory that we want. We start with Theory A given in (5.33) . It's just a complex scalar field with a $U(1)$ global symmetry $\phi \to e^{i\theta}\phi$. We'll deform this theory in the following way: we gauge the global symmetry and add a Chern-Simons term at level m. We end up with

$$S_A[a, \phi] = \int d^3x \; |\partial_\mu \phi - i a_\mu \phi|^2 - V(\phi) - \frac{m}{4\pi} \epsilon^{\mu\nu\rho} a_\mu \partial_\nu a_\rho \tag{5.36}$$

[47]For a more detailed discussion of these properties, see the *TASI Lectures on Solitons*.

But this is precisely our earlier effective action for the Laughlin state at filling fraction $\nu = 1/m$. In this context, the excitations of the field ϕ describe quasi-holes and quasi-particles in the theory, with fractional charge and statistics. The background gauge field of electromagnetism A_μ couples to the electron current which is

$$j^\mu = \frac{1}{2\pi}\epsilon^{\mu\nu\rho}\partial_\nu a_\rho$$

Now we can repeat this procedure for Theory B defined in (5.34). We again couple a $U(1)$ gauge field a_μ to the current which is now given by (5.35). We find

$$S_B[a,\alpha,\phi] = \int d^3x\ |\partial_\mu\tilde{\phi} - i\alpha_\mu\tilde{\phi}|^2 - V(\tilde{\phi}) + \frac{1}{2\pi}\epsilon^{\mu\nu\rho}a_\mu\partial_\nu\alpha_\rho - \frac{m}{4\pi}\epsilon^{\mu\nu\rho}a_\mu\partial_\nu a_\rho + \ldots$$

where the Maxwell term in (5.34) has been relegated to the ... in the expression above as it won't play an important role in what follows. Next we simply integrate out the gauge field a_μ in this Lagrangian. Because a_μ appears quadratically in the action, we can naive just replace it by its equation of motion which is

$$f_{\mu\nu} = \frac{1}{m}\tilde{f}_{\mu\nu}$$

Note, however, that we run into the same kind of issues that we saw in Section 5.2.2. This equation of motion is not consistent with the flux quantisation of both $f_{\mu\nu}$ and $\tilde{f}_{\mu\nu}$. This means that we should not take the resulting action too seriously when dealing with subtle topological issues, but hopefully it will capture the correct local physics. This action is:

$$S_B[\alpha,\phi] = \int d^3x\ |\partial_\mu\tilde{\phi} - i\alpha_\mu\tilde{\phi}|^2 - V(\tilde{\phi}) + \frac{1}{4\pi m}\epsilon^{\mu\nu\rho}\alpha_\mu\partial_\nu\alpha_\rho \ldots \tag{5.37}$$

This is the theory dual to (5.36). It is the dual description of the quantum Hall fluid. In the original theory (5.36), the elementary quanta ϕ are the quasi-particles while the vortices are the electrons. In the new description (5.37), the elementary quanta of $\tilde{\phi}$ are the electrons while the vortices are the quasi-particles.

There is one last step that is usually taken before we get to the final Ginzburg-Landau theory. The field $\tilde{\phi}$ in (5.37) has second order kinetic terms, which means that, upon quantisation, it will give rise to both particles and anti-particles. The particles are electrons (we will make this clearer below), while the anti-particles are holes. The existence of both particles and holes arises because both (5.36) and (5.37) describe physics around the quantum Hall state which, of course, is built upon a sea of electrons.

In contrast, in the Ginzburg-Landau approach to this problem it is more common to write down a field theory for electrons above the vacuum state. This is slightly odd because the resulting physics clearly requires a large number of electrons to be present but we can always insist upon this by including an appropriate chemical potential. We'll call the bosonic field that gives rise to electrons Φ. This field now has first order kinetic terms, reflecting the fact that there are no longer anti-particles. (Well, there are but they require around 10^{10} more energy than is available in quantum Hall system; this is condensed matter physics, not particle physics!). The resulting Lagrangian is

$$S = \int d^3x \;\; i\Phi^\dagger(\partial_0 - i\alpha_0 - i\mu)\Phi - \frac{1}{2m^\star}|\partial_i\Phi - i\alpha_i\Phi|^2 - V(\Phi) + \frac{1}{4\pi m}\epsilon^{\mu\nu\rho}\alpha_\mu\partial_\nu\alpha_\rho \quad (5.38)$$

with μ the promised chemical potential and $m^\star$ is the effective mass of the electron (and is not to be confused with the integer m). This is the proposed Chern-Simons Ginzberg-Landau description of the fractional quantum Hall effect. This Lagrangian was first written down by Zhang, Hansson and Kivelson and is sometime referred to as the ZHK theory[48].

Composite Bosons

We know from our previous discussion that the excitations of $\tilde{\phi}$ in (5.37) (or Φ in (5.38)) are supposed to describe the vortices of theory (5.36). Yet those vortices should carry the same quantum numbers as the original electrons. Let's first check that this makes sense.

Recall that the $\tilde{\phi}$ (or Φ) field is bosonic: it obeys commutation relations rather than anti-commutation relations. But, by the same arguments that we saw in Section 5.2.2, the presence of the Chern-Simons term will change the statistics of these excitations. In particular, if we work with the non-relativistic theory, the equation of motion for α_0 reads

$$\frac{1}{2\pi}\tilde{f}_{12} = -m\Phi^\dagger\Phi \quad (5.39)$$

Here $\Phi^\dagger\Phi$ is simply the particle density $n(\mathbf{x})$. This tells us that each particle in the quantum Hall fluid has $-m$ units of attached flux. By the usual Aharonov-Bohm arguments, these particles are bosonic if m is even and fermionic if m is odd. But that's exactly the right statistics for the "electrons" underlying the quantum Hall states.

[48]The original paper is S. C. Zhang, T. Hansson and S. Kivelson, "*Effective-Field-Theory Model for the Fractional Quantum Hall Effect*", Phys. Rev. Lett. **62**, 82 (1989) which can be downloaded here.

Let's briefly restrict to the case of m odd, so that the "electrons" are actual electrons. They can be thought of as bosons Φ attached to $-m$ flux units. Alternatively, the bosons Φ can be thought of as electrons attached to $+m$ units of flux. This object is referred to as a *composite boson*. Notice that it's very similar in spirit to the composite fermion that we met earlier. The difference is that we attach an odd number of fluxes to the electron to make a composite boson, while an even number of fluxes gives a composite fermion. In the next section, we'll see how to make a composite fermion in this language.

Off-Diagonal Long-Range Order

We took a rather round-about route to get to Lagrangian (5.38): we first looked at the most general description of a fractional quantum Hall effect, and subsequently dualised. However, it's possible to motivate (5.38) directly. In this short section, we briefly explain how.

The usual construction of a Ginzburg-Landau effective theory involves first identifying a symmetry which is broken. The symmetry breaking is then described by an appropriate local order parameter, and the effective theory is written in terms of this order parameter. If we want to do this for the quantum Hall fluid, we first need to figure out what this order parameter could possibly be.

We're going to take a hint from the theory of superfluidity where one works with an object called the *density matrix*. (Beware: this means something different than in previous courses on quantum mechanics and quantum information). There are two, equivalent, definitions of the density matrix. First, suppose that we have some many-body system with particles created by the operator $\Psi^\dagger(\mathbf{r})$. In a given state, we define the *density matrix* to be

$$\rho(\mathbf{r},\mathbf{r}') = \langle \Psi^\dagger(\mathbf{r})\Psi(\mathbf{r}')\rangle$$

Alternatively, there is also simple definition in the first quantised framework. Suppose that our system of N particles is described by the the wavefunction $\psi(\mathbf{x}_i)$. We focus on the position of just a single particle, say $\mathbf{x}_1 \equiv \mathbf{r}$ and the density matrix is constructed as

$$\rho(\mathbf{r},\mathbf{r}') = N\int \prod_{i=2}^{N} d\mathbf{x}_i\ \psi^\star(\mathbf{r},\mathbf{x}_2,\ldots,\mathbf{x}_N)\psi(\mathbf{r}',\mathbf{x}_2,\ldots,\mathbf{x}_N)$$

The definition of a *superfluid* state is that the density matrix exhibits *off-diagonal long range order*. This means that

$$\rho(\mathbf{r},\mathbf{r}') \to \rho_0 \quad \text{as} \quad |\mathbf{r}-\mathbf{r}'|\to\infty$$

Here ρ_0 is the density of the superfluid.

What does this have to do with our quantum Hall fluids? They certainly don't act like superfluids. And, indeed, you can check that quantum Hall fluids are *not* superfluids. If you compute the density matrix for the Laughlin wavefunction (3.3), you find

$$\rho(z,z') = N\int\prod_{i=2}^{N} d^2z_i\ \prod_i (z-z_i)^m(\bar{z}'-\bar{z}_i)^m \prod_{j<k}|z_j-z_k|^{2m}e^{-\sum_j |z_j|^2/2l_B^2}$$

This does not exhibit off-diagonal long-range order. The first two terms ensure that the phase fluctuates wildly and this results in exponential decay of the density matrix: $\rho(z,z') \sim e^{-|z-z'|^2}$.

However, one can construct an object which does exhibit off-diagonal long-range order. This is not apparent in the electrons, but instead in the composite bosons Φ. These operators are related to the electrons by the addition of $-m$ flux units,

$$\Phi^\dagger(z) = \Psi^\dagger(z)U^{-m} \tag{5.40}$$

where U is the operator which inserts a single unit of flux of the gauge field α_μ. It can be shown that this is the operator which exhibits off-diagonal long-range order in the quantum Hall state[49]

$$\langle\Phi^\dagger(z)\Phi(z')\rangle \to \rho_0 \quad \text{as} \quad |z-z'|\to\infty$$

Alternatively, if you're working with wavefunctions, you need to include a singular gauge transformation to implement the flux attachment.

Note that, usually in Ginzburg-Landau theories, one is interested in phases where the order parameter condensed. Indeed, if we follow through our duality transformations, the original theory (5.36) describes quantum Hall Hall physics when ϕ is a gapped excitation. (This is the phase $a > 0$ of Theory A in the previous section). But the particle-vortex duality tells us that the dual theory (5.37) should lie in the phase in which $\tilde{\phi}$ gets an expectation value. Equivalently, in the non-relativistic picture, Φ condenses.

[49]The first suggestion of long-range order in the Hall states was given by S. Girvin and A. H Macdonald, "*Off-Diagonal Long-Range Order, Oblique Confinement, and the Fractional Quantum Hall Effect*, Phys. Rev. Lett 58, 12 (1987). The refined, second-quantised arguments were given later by N. Read "*Order Parameter and Ginzburg-Landau Theory for the Fractional Quantum Hall Effect*", Phys. Rev. Lett. 62, 1 (1989).

This kind of thinking provided the original motivation for writing down the Ginzburg-Landau theory and, ultimately, to finding the link to Chern-Simons theories. However, the presence of the flux attachment in (5.40) means that Φ is not a local operator. This is one of the reasons why this approach misses some of the more subtle effects such as topological order.

Adding Background Gauge Fields

To explore more physics, we need to re-introduce the background gauge field A_μ into our effective Lagrangian. It's simple to re-do the integrating out with A_μ included; we find the effective Lagrangian

$$S = \int d^3x \left\{ i\Phi^\dagger(\partial_0 - i(\alpha_0 + A_0 + \mu))\Phi - \frac{1}{2m^\star}|\partial_i\Phi - i(\alpha_i + A_i)\Phi|^2 \right. \\ \left. -V(\Phi) + \frac{1}{4\pi m}\epsilon^{\mu\nu\rho}\alpha_\mu\partial_\nu\alpha_\rho \right\} \tag{5.41}$$

Because we're working with the non-relativistic theory, the excitations of Φ in the ground state should include all electrons in our system. Correspondingly, the gauge field A_μ should now include the background magnetic field that we apply to the system.

We've already seen that the Hall state is described when the Φ field condenses: $\langle\Phi^\dagger\Phi\rangle = n$, with n the density of electrons. But we pay an energy cost if there is a non-vanishing magnetic field B in the presence of such a condensate. This is the essence of the Meissner effect in a superconductor. However, our Hall fluid is not a superconductor. In this low-energy approach, it differs by the existence of the Chern-Simons gauge field α_μ which can turn on to cancel the magnetic field,

$$\alpha_i + A_i = 0 \quad \Rightarrow \quad \tilde{f}_{12} = -B$$

But we've already seen that the role of the Chern-Simons term is to bind the flux $\tilde{f}_{12}$ to the particle density $n(\mathbf{x})$ (5.39). We learn that

$$n(\mathbf{x}) = \frac{1}{2\pi m}B(\mathbf{x})$$

This is simply the statement that the theory is restricted to describe the lowest Landau level with filling fraction $\nu = 1/m$

We can also look at the vortices in this theory. These arise from the phase of Φ winding around the core of the vortex. The minimum vortex carries flux $\int d^2x\ \tilde{f}_{12} =$

$\pm 2\pi$. From the flux attachment (5.39), we see that they carry charge $e^\star = \pm 1/m$. This is as expected from our general arguments of particle-vortex duality: the vortices in the ZHK theory should correspond to the fundamental excitations of the original theory (5.36): these are the quasi-holes and quasi-particles.

So far, we've seen that this dual formalism can reproduce many of the results that we saw earlier. However, the theory (5.41) provides a framework to compute much more detailed response properties of the quantum Hall fluid. For most of these, it is not enough to consider just the classical theory as we've done above. One should take into account the quantum fluctuations of the Chern-Simons field, as well as the Coulomb interactions between electrons which we've buried in the potential. We won't describe any of this here[50].

5.3.3 Composite Fermions and the Half-Filled Landau Level

We can also use this Chern-Simons approach to make contact with the composite fermion picture that we met in Section 3. Recall that the basic idea was to attach an *even* number of vortices to each electron. In the language of Section 3, these vortices were simply zeros of the wavefunction, with holomorphicity ensuring that each zero is accompanied by a 2π winding of the phase. In the present language, we can think of the vortex attachment as flux attachment. Adding an even number of fluxes to an electron doesn't change the statistics. The resulting object is the composite fermion.

As we saw in Section 3.3.3, one of the most interesting predictions of the composite fermion picture arises at $\nu = 1/2$ where one finds a compressible fermi-liquid-type state. We can write down an effective action for the half-filled Landau level as follows,

$$S = \int d^3x \quad \left\{ i\psi^\dagger(\partial_0 - i(\alpha_0 + A_0 + \mu)\psi - \frac{1}{2m^\star}|\partial_i\psi - i(\alpha_i + A_i)\psi|^2 \right. \\ \left. + \frac{1}{2}\frac{1}{4\pi}\epsilon^{\mu\nu\rho}\alpha_\mu\partial_\nu\alpha_\rho + \frac{1}{2}\int d^2x' \ \psi^\dagger(\mathbf{x})\psi(x)V(\mathbf{x}-\mathbf{x}')\psi^\dagger(\mathbf{x}')\psi(\mathbf{x}') \right\} \tag{5.42}$$

Here ψ is to be quantised as a fermion, obeying anti-commutation relations. We have also explicitly written the potential between electrons, with $V(\mathbf{x})$ usually taken to the be the Coulomb potential. Note that the Chern-Simons term has coefficient $1/2$, as befits a theory at half-filling.

[50]For a nice review article, see Shou Cheng Zhang, "*The Chern-Simons-Landau-Ginzburg Theory of the Fractional Quantum Hall Effect*, Int. Jour. Mod. Phys. **B6** (1992).

The action (5.42) is the starting point for the Halperin-Lee-Read theory of the half-filled Landau level. The basic idea is that an external magnetic field B can be screened by the emergent gauge field $\tilde{f}_{12}$, leaving the fermions free to fill up a Fermi sea. However, the fluctuations of the Chern-Simons gauge field mean that the resulting properties of this metal are different from the usual Fermi-liquid theory. It is, perhaps, the simplest example of a "non-Fermi liquid". Many detailed calculations of properties of this state can be performed and successfully compared to experiment. We won't describe any of this here[51].

Half-Filled or Half-Empty?

While the HLR theory (5.42) can claim many successes, there remains one issue that is poorly understood. When a Landau level is half full, it is also half empty. One would expect that the resulting theory would then exhibit a symmetry exchanging particles and holes. But the action (5.42) does not exhibit any such symmetry.

There are a number of logical possibilities. The first is that, despite appearances, the theory (3.43) does secretly preserve particle-hole symmetry. The second possibility is that this symmetry is spontaneously broken at $\nu = 1/2$ and there are actually two possible states. (This turns out to be true at $\nu = 5/2$ where the Pfaffian state we've already met has a brother, known as the anti-Pfaffian state).

Here we will focus on a third possibility: that the theory (5.42) is not quite correct. An alternative theory was suggested by Son who proposed that the composite fermion at $\nu = 1/2$ should be rightly viewed as a two-component Dirac fermion[52].

The heart of Son's proposal is a new duality that can be thought of as a fermionic version of the particle-vortex duality that we met in Section 5.3.1. Here we first describe this duality. In the process of explaining how it works, we will see the connection to the half-filled Landau level.

Theory A: The Dirac Fermion

Our first theory consists of a single Dirac fermion ψ in $d = 2 + 1$ dimensions

$$S_A = \int d^3x \; i\bar{\psi}(\partial\!\!\!/ - i A\!\!\!/)\psi + \ldots \tag{5.43}$$

[51]Details can be found in the original paper by Halperin, Lee and Read, "*Theory of the half-filled Landau level*", Phys. Rev. B 47, 7312 (1993), and in the nice review by Steve Simon, "*The Chern-Simons Fermi Liquid Description of Fractional Quantum Hall States*", cond-mat/9812186.

[52]Son's original paper is "*Is the Composite Fermion a Dirac Particle?*", Phys. Rev. **X5**, 031027 (2015), arXiv:1502.03446.

In $d = 2+1$ dimensions, the representation of the Clifford algebra $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}$ has dimension 2. The gamma matrices can be written in terms of the Pauli matrices, with a useful representation given by

$$\gamma^0 = i\sigma^2 \quad , \quad \gamma^1 = \sigma^1 \quad , \quad \gamma^2 = \sigma^3$$

Correspondingly, the Dirac spinor ψ is a two-component object with complex components. As usual, $\bar{\psi} = \psi^\dagger \gamma^0$. (See the lectures on *Quantum Field Theory* for more information about the construction of spinors). Quantising the Dirac spinor in $d = 2+1$ dimensions gives rise to spin-up particles and spin-down anti-particles.

Theory A has a global $U(1)$ symmetry with current

$$J^\mu = \bar{\psi}\gamma^\mu\psi \tag{5.44}$$

In the action (5.43), we've coupled this to a background electromagnetic gauge field A_μ.

Theory B: QED$_3$

The second theory also consists of a single Dirac fermion, $\tilde{\psi}$, this time coupled to a dynamical $U(1)$ gauge field α_μ.

$$S_B = \int d^3x \; i\bar{\tilde{\psi}}(\partial\!\!\!/ - 2i\,\alpha\!\!\!/)\tilde{\psi} + \frac{1}{2\pi}\epsilon^{\mu\nu\rho}\alpha_\mu\partial_\nu A_\rho + \ldots \tag{5.45}$$

This is essentially QED in $d = 2+1$ dimensions. However, there is one crucial subtlety: $\tilde{\psi}$ carries charge 2 under this gauge field, not charge 1. To avoid rescaling of the gauge field, we should accompany this charge with the statement that the fluxes of α remain canonically normalised

$$\frac{1}{2\pi}\int_{\mathbf{S}^2} \tilde{f}_{12} \in \mathbf{Z}$$

The charge 2 is crucial for this theory to make sense. If the fermion $\tilde{\psi}$ had charge 1 then the theory wouldn't make sense: it suffers from a discrete gauge anomaly, usually referred to as a *parity anomaly* in this context. However, with charge 2 this is avoided[53]
.

[53]This is actually a bit too quick. A more careful analysis was given by T. Senthil, N. Seiberg, E. Witten and C. Wang in "*A Duality Web in 2+1 Dimensions and Condensed Matter Physics*", ArXiv:1606.01989.

The theory (5.45) has a $U(1)$ symmetry with the kind of current that is, by now, familiar

$$J^\mu = \frac{1}{2\pi}\epsilon^{\mu\nu\rho}\partial_\nu \alpha_\rho$$

This is to be identified with the current (5.44) of Theory A.

Half-Filling in the Two Theories

Let's start with Theory A and turn on a background magnetic field B. The Dirac fermions form Landau levels. However, because of the relativistic dispersion relation, these Landau levels are somewhat different from those we met in Section 1. A simple generalisation of these calculations shows that the Landau levels have energy

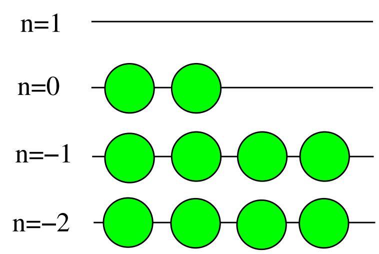


Figure 44: The view from Theory A

$$E_n^2 = 2B|n| \quad n \in \mathbf{Z}$$

Note, in particular, that there is a zero energy $n = 0$ Landau level. This arises because the zero-point energy $\frac{1}{2}\hbar\omega_B$ seen in the non-relativistic Landau levels is exactly compensated by the Zeeman splitting.

In the Dirac sea picture, we can think of filling the negative energy Landau levels, which we label with $n < 0$. However, if we restrict to zero density then the $n = 0$ Landau level is necessarily at half-filling. This is shown in the picture. In the absence of any interactions there is a large degeneracy. We rely on the interactions, captured by the ... in (5.43), to resolve this degeneracy. In this way, the Dirac fermion in a magnetic field automatically sits at half filling. Note that this picture is, by construction, symmetric under interchange of particles and holes.

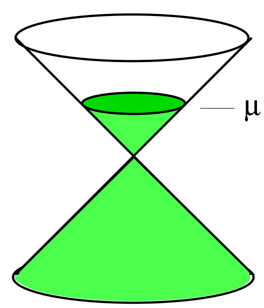


Figure 45: and from Theory B

Let's now see what this same picture looks like in Theory B. The background magnetic field contributes a term $\frac{1}{2\pi}B\alpha_0$ to the action (5.45). This is a background charge density, $\tilde{n} = \frac{1}{2}(B/2\pi)$, where the factor of $1/2$ can be traced to to the charge 2 carried by the fermion. This means that the fermions in QED_3 pile up to form a Fermi sea, with chemical potential μ set by the background magnetic field. This is shown in the figure to the right.

This is the new proposed dual of the half-filled Landau level. We see that there is no hint of the magnetic field in the dual picture. Instead we get a Fermi surface which, just as in the HLR theory (5.42), is coupled to a fluctuating gauge field. However, in this new proposal this gauge field no longer has a Chern-Simons coupling.

It turns out that many, if not all, of the successful predictions of the HLR theory (5.42) also hold for QED_3 (5.45). The difference between the theories two turns out to be rather subtle: the relativistic electrons in QED_3 pick up an extra factor of Berry phase π as they are transported around the Fermi surface. At the time of writing, there is an ongoing effort to determine whether this phase can be observed experimentally see which of these two theories is correct.

5.4 Non-Abelian Chern-Simons Theories

So far we have discussed the effective theories only for Abelian quantum Hall states. As we have seen, these are described by Chern-Simons theories involving emergent $U(1)$ gauge fields. Given this, it seems plausible that the effective field theories for non-Abelian quantum Hall states involve emergent non-Abelian Chern-Simons gauge fields. This is indeed the case. Here we sketch some of the novel properties associated to non-Abelian Chern-Simons terms.

5.4.1 Introducing Non-Abelian Chern-Simons Theories

We start by describing the basics of non-Abelian Chern-Simons theories. Everything we say will hold for arbitrary gauge group G, but we will focus on $G = SU(N)$. For the most prominent applications to quantum Hall physics, $G = SU(2)$ will suffice. We work with Hermitian gauge connections a_μ, valued in the Lie algebra. The associated field strength is

$$f_{\mu\nu} = \partial_\mu a_\nu - \partial_\nu a_\mu - i[a_\mu, a_\nu]$$

We take the basis generators in the fundamental representation with normalisation $\mathrm{tr}(T^a T^b) = \frac{1}{2}\delta^{ab}$. With this choice, the Yang-Mills action takes the familiar form

$$S_{YM} = -\frac{1}{2g^2}\int d^3x \ \mathrm{tr}\, f^{\mu\nu} f_{\mu\nu}$$

However, just as we saw for the Abelian gauge fields, we are not interested in the Yang-Mills action. Instead, there is an alternative action that we can write down in $d = 2+1$ dimensions. This is the non-Abelian Chern-Simons action

$$S_{CS} = \frac{k}{4\pi}\int d^3x \ \epsilon^{\mu\nu\rho}\mathrm{tr}\left(a_\mu \partial_\nu a_\rho - \frac{2i}{3} a_\mu a_\nu a_\rho\right) \tag{5.46}$$

Chern-Simons theories with gauge group G and level k are sometimes denoted as G_k.

Our first goal is to understand some simple properties of this theory. The equation of motion is

$$f_{\mu\nu} = 0$$

This is deceptively simple! Yet, as we will see, many of the subtleties arise from the interesting solutions to this equation and its generalisations. Indeed, we've already seen our first hint that this equation has interesting solutions when we looked at Abelian Chern-Simons theories on the torus in Section 5.2.3.

Let's start by seeing how the Chern-Simons action fares under a gauge transformation. The gauge potential transforms as

$$a_\mu \to g^{-1} a_\mu g + i g^{-1} \partial_\mu g$$

with $g \in SU(N)$. The field strength transforms as $f_{\mu\nu} \to g^{-1} f_{\mu\nu} g$. A simple calculation shows that the Chern-Simons action changes as

$$S_{CS} \to S_{CS} + \frac{k}{4\pi} \int d^3x \left\{ \epsilon^{\mu\nu\rho} \partial_\nu \mathrm{tr}\,(\partial_\mu g\, g^{-1} a_\rho) + \frac{1}{3} \epsilon^{\mu\nu\rho} \mathrm{tr}\,(g^{-1}\partial_\mu g\, g^{-1}\partial_\nu g\, g^{-1}\partial_\rho g) \right\}$$

The first term is a total derivative. The same kind of term arose in Abelian Chern-Simons theories. It will have an interesting role to play on manifolds with boundaries.

For now, our interest lies in the second term. This is novel to non-Abelian gauge theories and has a beautiful interpretation. To see this, consider our theory on Euclidean $\mathbf{S}^3$ (or on $\mathbf{R}^3$ with the requirement that gauge transformations asymptote to the same value at infinity). Then the gauge transformations can "wind" around spacetime. This follows from the homotopy group $\Pi_3(SU(N)) \cong \mathbf{Z}$. The winding is counted by the function

$$w(g) = \frac{1}{24\pi^2} \int d^3x\ \epsilon^{\mu\nu\rho} \mathrm{tr}\,(g^{-1}\partial_\mu g\, g^{-1}\partial_\nu g\, g^{-1}\partial_\rho g) \quad \in \mathbf{Z} \tag{5.47}$$

We recognise this as the final term that appears in the variation of the Chern-Simons action. This means that the Chern-Simons action is not invariant under these large gauge transformations; it changes as

$$S_{CS} \to S_{CS} + \frac{k}{12\pi} 24\pi^2 w(g) = S_{CS} + 2\pi k w(g)$$

However, just as we saw earlier, we need not insist that the Chern-Simons action is invariant. We need only insist that the exponential that appears in the path integral, $e^{iS_{CS}}$ is invariant. We see that this holds providing

$$k \in \mathbf{Z}$$

This is the same quantisation that we saw for the Abelian theory, although this requirement arises in a more direct fashion for the non-Abelian theory. (Note that we're using the convention $e = \hbar = 1$; if we put these back in, we find $\hbar k/e^2 \in \mathbf{Z}$).

Chern-Simons Term as a Boundary Term

There is one other basic property of the Chern-Simons term that is useful to know. Consider a theory in $d = 3+1$ dimensions. A natural quantity is the Pontryagin density $\epsilon^{\mu\nu\rho\sigma}\mathrm{Tr}(f_{\mu\nu}f_{\rho\sigma})$. It's not hard to show that this is a total derivative,

$$\epsilon^{\mu\nu\rho\sigma}\mathrm{tr}(f_{\mu\nu}f_{\rho\sigma}) = 4\epsilon^{\mu\nu\rho\sigma}\partial_\mu\,\mathrm{tr}\,\left(a_\mu\partial_\rho a_\sigma - \frac{2i}{3}a_\nu a_\rho a_\sigma\right)$$

The object in brackets is precisely the Chern-Simons term.

5.4.2 Canonical Quantisation and Topological Order

Let's now quantise the Chern-Simons theory (5.46). Here, and also in Section 5.4.4, we explain how to do this. However, both sections will be rather schematic, often stating results rather than deriving them[54]. We'll consider the theory on a manifold $\mathbf{R}\times\Sigma$ where $\mathbf{R}$ is time and Σ is a spatial manifold which we'll take to be compact. Mostly in what follows we'll be interested in $\Sigma = \mathbf{S}^2$ and $\Sigma = \mathbf{T}^2$, but we'll also present results for more general manifolds. The action (5.46) can then be written as

$$S_{CS} = \frac{k}{4\pi}\int dt\int_\Sigma d^2x\ \mathrm{tr}\,\left(\epsilon^{ij}a_i\frac{\partial}{\partial t}a_j + a_0 f_{12}\right) \tag{5.48}$$

This is crying out to be quantised in $a_0 = 0$ gauge. Here, the dynamical degrees of freedom a_i obey the commutation relations

$$[a_i^a(\mathbf{x}), a_j^b(\mathbf{y})] = \frac{2\pi i}{k}\,\epsilon_{ij}\,\delta^{ab}\,\delta^2(\mathbf{x}-\mathbf{y}) \tag{5.49}$$

Subject to the constraint

$$f_{12} = 0 \tag{5.50}$$

As always with a gauge theory, there are two ways to proceed. We could either quantise and then impose the constraint. Or we could impose the constraint classically and quantise the resulting degrees of freedom. Here, we start by describing the latter approach.

[54]There is a long and detailed literature on this material, starting with Edward Witten's Fields medal winning work, "*Quantum Field Theory and the Jones Polynomial*", Comm. Math. Phys. Volume 121, Number 3, 351 (1989).

We're looking for solutions to (5.50) on the background Σ. This is the problem of finding *flat connections* on Σ and has been well studied in the mathematical literature. We offer only a sketch of the solution. We already saw in Section 5.2.3 how to do this for Abelian Chern-Simons theories on a torus: the solutions are parameterised by the holonomies of a_i around the cycles of the torus. The same is roughly true here. For gauge group $SU(N)$, there are $N^2 - 1$ such holonomies for each cycle, but we also need to identify connections that are related by gauge transformations. The upshot is that the moduli space $\mathcal{M}$ of flat connections has dimension $(2g-2)(N^2-1)$ where g is the genus Σ.

Usually in classical mechanics, we would view the space of solutions to the constraint – such as $\mathcal{M}$ – as the configuration space of the system. But that's not correct in the present context. Because we started with a first order action (5.48), the a_i describe both positions and momenta of the system. This means that $\mathcal{M}$ is the *phase space*. Now, importantly, it turns out that the moduli space $\mathcal{M}$ is compact (admittedly with some singularities that have to be dealt with). So we're in the slightly unusual situation of having a compact phase space. When you quantise you (very roughly) parcel the phase space up into chunks of area $\hbar$. Each of these chunks corresponds to a different state in the quantum Hilbert space. This means that when you have a compact phase space, you will get a finite number of states. Of course, this is precisely what we saw for the $U(1)$ Chern-Simons theory on a torus in Section 5.2.3. What we're seeing here is just a fancy way of saying the same thing.

So the question we need to answer is: what is the dimension of the Hilbert space $\mathcal{H}$ that you get from quantising $SU(N)$ Chern-Simons theory on a manifold Σ?

When $\Sigma = \mathbf{S}^2$, the answer is easy. There are no flat connections on $\mathbf{S}^2$ and the quantisation is trivial. There is just a unique state: $\dim(\mathcal{H}) = 1$. In Section 5.4.4, we'll see how we can endow this situation with something a little more interesting.

When Σ has more interesting topology, the quantisation of G_k leads to a more interesting Hilbert space. When $G = SU(2)$, it turns out that the dimension of the Hilbert space for $g \geq 1$ is[55]

$$\dim(\mathcal{H}) = \left(\frac{k+2}{2}\right)^{g-1} \sum_{j=0}^{k} \left(\sin\frac{(j+1)\pi}{k+2}\right)^{2(g-1)} \tag{5.51}$$

[55]This formula was first derived using a connection to conformal field theory. We will touch on this in Section 6. The original paper is by Eric Verlinde, "*Fusion Rules and Modular Invariance in 2d Conformal Field Theories*", Nucl. Phys. **B300**, 360 (1988). It is sometimes referred to the Verlinde formula.

Note that for $\Sigma = \mathbf{T}^2$, which has $g = 1$, this is simply $\dim(\mathcal{H}) = k + 1$. It's not obvious, but nonetheless true, that the formula above gives an integer for all g. There is a generalisation of this formula for general gauge group which involves various group theoretic factors such as sums over weights.

Finally, note that the dimension of the Hilbert space can be computed directly within the path integral. One simply needs to compute the partition function on the manifold $\mathbf{S}^1 \times \Sigma$,

$$Z = \int \mathcal{D}a \ \exp\left[\frac{ik}{4\pi}\int_{\mathbf{S}^1\times\Sigma} d^3x \ \epsilon^{\mu\nu\rho}\mathrm{tr}\left(a_\mu\partial_\nu a_\rho - \frac{2i}{3}a_\mu a_\nu a_\rho\right)\right] = \dim(\mathcal{H})$$

This provides a more direct way of computing the dimensions (5.51) of the Hilbert spaces[56].

The discussion above has been rather brief. It turns out that the best way to derive these results is to map the problem into a $d = 1+1$ conformal field theory known as the WZW model. Indeed, one of the most surprising results in this subject is that there is a deep connection between the states of the Chern-Simons theory and objects known as conformal blocks in the WZW model. We'll comment briefly on this in Section 6.

5.4.3 Wilson Lines

So far we've only discussed the pure Chern-Simons. Now we want to introduce new degrees of freedom that are charged under the gauge field. These will play the role of non-Abelian anyons in the theory.

In the case of Abelian Chern-Simons theories, we could introduce quasi-holes by simply adding a background current to the Lagrangian. In the non-Abelian case, we need to be a little more careful. A current J^μ couples to the gauge field as,

$$\int d^3x \ \mathrm{tr}\,(a_\mu J^\mu)$$

But now the current must transform under the gauge group. This means that we can't just stipulate some fixed background current because that wouldn't be gauge invariant. Instead, even if the charged particle is stationary, the current must include some dynamical degrees of freedom. These describe the internal orientation of the particle within the gauge group. In the language of QCD, they are the "colour" degrees

[56] This calculation was described in M. Blau and G. Thompson, "*Derivation of the Verlinde Formula from Chern-Simons Theory and the G/G Model*", Nucl. Phys. **408**, 345 (1993), hep-th/9305010 where clear statements of the generalisation to other groups can be found.

of freedom of each quark and we'll retain this language here. In general, these colour degrees of freedom span some finite dimensional Hilbert space. For example, if we have an object transforming in the fundamental representation of $SU(N)$, then it will have an N-dimensional internal Hilbert space .

In this section we'll see how to describe these colour degrees of freedom for each particle. Usually this is not done. Instead, one can work in a description where the colour degrees of freedom are integrated out in the path integral, leaving behind an object called a *Wilson line*. The purpose of this Section is really to explain where these Wilson lines come from. In Section 5.4.4, we will return to Chern-Simons theories and study their properties in the presence of these external sources.

Classically, we view each particle that is charged under the $SU(N)$ gauge field as carrying an internal N-component complex vector with components w_γ, $\gamma = 1, \ldots, N$. This vector has some special properties. First, it has a fixed length

$$w^\dagger w = \kappa \tag{5.52}$$

Second, we identify vectors which differ only by a phase: $w_\gamma \sim e^{i\theta} w_\gamma$. This means that the vectors parameterise the projective space $\mathbf{CP}^{N-1}$.

Let's ignore the coupling of to the gauge field a_μ for now. The dynamics of these vectors is described by introducing an auxiliary $U(1)$ gauge field α which lives on the worldline of the particle. The action is

$$S_w = \int dt \ \left(i w^\dagger \mathcal{D}_t w - \kappa \alpha \right) \tag{5.53}$$

where $\mathcal{D}_t = \partial_t w - i\alpha w$. The purpose of this gauge field is two-fold. Firstly, we have a gauge symmetry which identifies $w \to e^{i\theta(t)} w$. This means that two vectors which differ only by a phase are physically equivalent, just as we wanted. Second, the equation of motion for α is precisely the constraint equation (5.52). The net result is that w_γ indeed parameterise $\mathbf{CP}^{N-1}$.

Note, however, that our action is first order, rather than second order. This means that $\mathbf{CP}^{N-1}$ is the phase space of the colour vector rather than the configuration space. But this too is what we want: whenever we quantise a compact phase space, we end up with a finite dimensional Hilbert space.

Finally, we can couple the colour degree of freedom to the Chern-Simons gauge field. If the particle is stationary at some fixed position $\mathbf{x} = \mathbf{X}$, then the action is

$$S_w = \int dt \ \left(i w^\dagger \mathcal{D}_t w - \kappa \alpha - w^\dagger a_0(t) w \right)$$

where $a_0(t) = a_0(t, \mathbf{x} = \mathbf{X})$ is the Chern-Simons gauge field at the location of the particle. The equation of motion for w is then

$$i \frac{dw}{dt} = a_0(t) w$$

In other words, the Chern-Simons gauge field tells this colour vector how to precess.

Quantising the Colour Degree of Freedom

It's straightforward to quantise this system. Let's start with the unconstrained variables w_γ which obey the commutation relations,

$$[w_\gamma, w^\dagger_{\gamma'}] = \delta_{\gamma\gamma'} \tag{5.54}$$

We define a "ground state" $|0\rangle$ such that $w_\gamma |0\rangle = 0$ for all $\gamma = 1, \ldots, N$. A general state in the Hilbert space then takes the form

$$|\gamma_1 \ldots \gamma_n\rangle = w^\dagger_{\gamma_1} \ldots w^\dagger_{\gamma_n} |0\rangle$$

However, we also need to take into account the constraint (5.52) which, in this context, arises from the worldline gauge field α. In the quantum theory, there is a normal ordering ambiguity in defining this constraint. The symmetric choice is to take the charge operator

$$Q = \frac{1}{2}(w^\dagger_\gamma w_\gamma + w_\gamma w^\dagger_\gamma) \tag{5.55}$$

and to impose the constraint

$$Q = \kappa \tag{5.56}$$

The spectrum of Q is quantised which means that the theory only makes sense if κ is also quantised. In fact, the $\kappa\alpha$ term in (5.53) is the one-dimensional analog of the 3d Chern-Simons term. (In particular, it is gauge invariant only up to a total derivative). The quantisation that we're seeing here is very similar to the kind of quantisations that we saw in the 3d case.

However, the normal ordering implicit in the symmetric choice of Q in (5.55) gives rise to a shift in the spectrum. For N even, Q takes integer values; for N odd, Q takes half-integer values. It will prove useful to introduce the shifted Chern-Simons coefficient,

$$\kappa_{\text{eff}} = \kappa - \frac{N}{2} \tag{5.57}$$

The quantisation condition then reads $\kappa_{\text{eff}} \in \mathbf{Z}^+$.

The constraint (5.56) now restricts the theory to a finite dimensional Hilbert space, as expected from the quantisation of a compact phase space $\mathbf{CP}^{N-1}$. Moreover, for each value of κ_{eff}, the Hilbert space inherits an action under the $SU(N)$ global symmetry. Let us look at some examples:

- $\kappa_{\text{eff}} = 0$: The Hilbert space consists of a single state, $|0\rangle$. This is equivalent to putting a particle in the trivial representation of the gauge group.
- $\kappa_{\text{eff}} = 1$: The Hilbert space consists of N states, $w_\gamma^\dagger|0\rangle$. This describes a particle transforming in the fundamental representation of the $SU(N)$ gauge group.
- $\kappa_{\text{eff}} = 2$: The Hilbert space consists of $\frac{1}{2}N(N+1)$ states, $w_\gamma^\dagger w_{\gamma'}^\dagger|0\rangle$, transforming in the symmetric representation of the gauge group.

By increasing the value of κ_{eff} in integer amounts, it is clear that we can build all symmetric representations of $SU(N)$ in this manner. If we were to replace the commutators in (5.54) with anti-commutators, $\{w_\gamma, w_{\gamma'}^\dagger\} = \delta_{\gamma\gamma'}$, then it's easy to convince yourself that we will end up with particles in the anti-symmetric representations of $SU(N)$.

The Path Integral

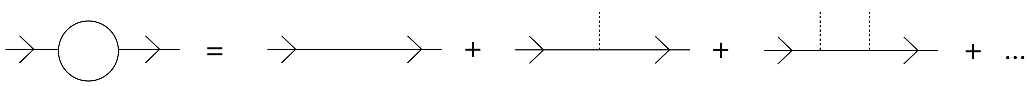
Let's now see what happens if we compute the path integral. For now, we will fix the Chern-Simons field $a_0(t)$ and consider only the integral over w and the worldline gauge field α. Subsequently, we'll also integrate over a_μ.

The path integral is reasonably straightforward to compute. One has to be a little careful with the vacuum bubbles whose effect is to implement the shift (5.57) from the path integral perspective. Let's suppose that we want to compute in the theory with $\kappa_{\text{eff}} = 1$, so we're looking at objects in the $\mathbf{N}$ representation of $SU(N)$. It's not hard to see that the path integral over α causes the partition function to vanish unless we put in two insertions of w. We should therefore compute

$$W[a_0] = \int \mathcal{D}\alpha\mathcal{D}w\mathcal{D}w^\dagger \; e^{iS_w(w,\alpha;a_0)} w_\gamma(t=\infty)w_\gamma^\dagger(t=-\infty)$$

Note that we've called the partition function W as opposed to its canonical name Z. We'll see the reason for this below. The insertion at $t = -\infty$ is simply placing the particle in some particular internal state and the partition function measures the amplitude that it remains in that state at $t = +\infty$

Having taken this into account, we next perform the path integral over w and $w^\dagger$. This is tantamount to summing a series of diagrams like this:



where the straight lines are propagators for w_γ which are simply $\theta(t_1 - t_2)\delta_{\gamma\gamma'}$, while the dotted lines represent insertions of the gauge fields. It's straightforward to sum these. The final result is something very simple:

$$W[a_0] = \mathrm{Tr}\,\mathcal{P}\exp\left(i\int dt\; a_0(t)\right) \tag{5.58}$$

Here $\mathcal{P}$ stands for path ordering which, since our particles are static, is the same thing as time ordering. The trace is evaluated in the fundamental representation. This is the *Wilson line*. It is a classical function of the gauge field $a_0(t)$. However, as we've seen above, it should really be thought of as a quantum object, arising from integrating out the colour degrees of freedom of a particle.

We can also generalise this construction to other symmetric representations; you simply need to insert κ_{eff} factors of $w^\dagger$ at time $t = -\infty$ and a further κ_{eff} factors of w at $t = +\infty$. The end result is a Wilson line, with the trace evaluated in the $\kappa_{\text{eff}}^{\text{th}}$ symmetric representation.

5.4.4 Chern-Simons Theory with Wilson Lines

Let's now consider non-Abelian Chern-Simons theory with the insertion of some number of Wilson lines. Suppose that we insert n Wilson lines, each in a representation R_i and sitting at position $\mathbf{X}_i$. For simplicity, we'll consider the theory on $\mathbf{R}\times\mathbf{S}^2$ where, previously, the theory had just a single state. Now we quantise in the presence of these Wilson lines. This will give a new Hilbert space that we'll denote $\mathcal{H}_{i_1\ldots i_n}$ with the labels denoting both position and representation of the Wilson lines. The first question that we want to ask is: what is the dimension of this new Hilbert space?

The constraint equation in the presence of Wilson lines reads

$$\frac{k}{2\pi}f^a_{12}(\mathbf{x}) = \sum_{i=1}^{n}\delta^2(\mathbf{x}-\mathbf{X}_i)\,w^{(i)\dagger}T^a w^{(i)} \tag{5.59}$$

with $w^{(i)}$ the colour degrees of freedom that we met in the previous section. These carry the information about the representation R_i carried by the Wilson line

Let's start by looking at the limit $k \to \infty$. This is the weak coupling limit of the Chern-Simons theory (strictly, we need $k \gg N$) so we expect a classical analysis to be valid. However, we'll retain one element of the quantum theory: the Dirac quantisation of flux (5.19), now applied to each component f_{12}^a individually. But, with k very large, we see that it's impossible to reconcile Dirac quantisation with any non-trivial charge on the right-hand side. This means that the only way we can solve (5.59) is if the charges on the right-hand side can somehow add up to zero. In the language of group theory, this means that we take need to decompose the tensor product of the representations R_i into irreducible representations. We only get solutions to (5.59) only if singlets appear in this decomposition. We write

$$\otimes_{i=1}^n R_i = \mathbf{1}^p \oplus \ldots$$

where p is the number of singlets $\mathbf{1}$ appearing in the decomposition and $\ldots$ are all the non-singlet representations. Each of these different decompositions gives rise to a different state in the Hilbert space $\mathcal{H}_{i_1 \ldots i_n}$. In the weak coupling limit, we then have

$$\lim_{k\to\infty} \dim(\mathcal{H}_{i_1\ldots i_n}) = p$$

Typically, when we have a large number n of Wilson lines, there will be several different ways to make singlets so $p \geq 2$.

For finite k when quantum effects become more important, one finds that

$$\dim(\mathcal{H}_{i_1\ldots i_n}) \leq p$$

The possible reduction of the number of states arises in an intuitive fashion through screening. At finite k, new solutions to (5.59) exist in which the integrated flux is non-zero. But we should sum over flux sectors in the path integral which means that these states become indistinguishable from the vacuum. This not only cuts down the dimension of the Hilbert space, but reduces the kinds of representations that we can insert to begin with. Let's illustrate this idea with some simple examples:

An Example: $SU(2)_k$

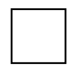
For $G = SU(2)$, representations are labelled by the spin s. Classically, of course, s can take any half-integer value. There is no bound on how large the spin can be. However, at finite k the spin is bounded by

$$0 \leq s \leq \frac{k}{2} \tag{5.60}$$

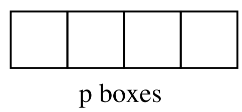
The insertion of any Wilson line with spin $s > k/2$ can be screened by flux so that it is equivalent to spin $|s-k|$.

Another Example: $SU(N)_k$.

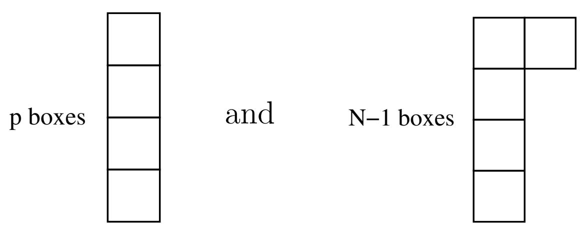
Let's first recall some $SU(N)$ group theory. Irreducible representations can be characterised by a Young tableau with rows of length $l_1 \geq l_2 \geq \ldots \geq l_{N-1} \geq 0$. In this notation, the fundamental representation $\mathbf{N}$ is simply a single box



The p^{th} symmetric representation is a row of boxes



The anti-symmetric representation is a column of p boxes, while the adjoint is a full column plus an extra guy stuck on the top,



In particular, the anti-fundamental representation $\bar{\mathbf{N}}$ is the same as the $(N-1)^{\text{th}}$ anti-symmetric representation.

The non-trivial Wilson lines at level k are simply those with $l_1 \leq k$. This means, in particular, that we can only have symmetric representations up to the k^{th} power of the fundamental. (This agrees with our result (5.60) for $SU(2)$). However, all anti-symmetric representations are allowed. Most importantly, there are only a finite number of representations at any finite k.

Fusion Revisited

Having specified the allowed representations, let's now return to the dimension of the Hilbert space $\mathcal{H}_{i_1 \ldots i_n}$. For two Wilson lines, the Hilbert space has dimension 1 if $R_1 = \bar{R}_2$, so that their tensor product can form a singlet. The first non-trivial example arises with the insertion of three Wilson lines with representations as R_i, R_j and $\bar{R}_k$. We'll denote the dimension of the Hilbert space as

$$\dim(\mathcal{H}_{ij\bar{k}}) = N_{ij}^k$$

As we described above, in the classical limit N_{ij}^k is the number of times that R_k appears in the tensor product of $R_i \otimes R_j$. However, it too can receive quantum corrections and, in general, N_{ij}^k will be less than its classical value.

There is a well-developed machinery to compute the numbers N_{ij}^k in Chern-Simons theories. This involves replacing the tensor product of representations $\otimes$ with a modified operation called *fusion*. We will denote the fusion of two representations as $\star$. The number N_{ij}^k is now the number of times that R_k appears in the fusion product of $R_i \star R_j$.

From knowledge of the N_{ij}^k, we can compute the dimension of the general Hilbert space $\mathcal{H}_{i_1 \ldots i_n}$. It is given by

$$\dim(\mathcal{H}_{i_1 \ldots i_n}) = \sum_{j_1,\ldots,j_{n-2}} N_{i_1 i_2}^{j_1} N_{j_1 i_3}^{j_2} \ldots N_{j_{n-2} i_n}^{j_{n-1}}$$

We've seen all of this before. This is the formal structure of *fusion* that underlies the theory of non-Abelian anyons that we described in Section 4.3. The formula above is the same as (4.21). In general, the Hilbert space of Wilson lines in Chern-Simons theory provides a concrete realisation of the somewhat abstract fusion rules.

The fusion rules for Wilson lines in Chern-Simons theories are related to the representation theory of Kac-Moody algebras. We won't explain where these rules come from. Instead, we will just present the results[57].

Fusion Rules for $SU(2)_k$

The representations of $SU(2)$ are labelled by the spin s or the dimension $\mathbf{d} = 2s + 1$. The tensor product between two representations follows from the familiar Clebsh-Gordon decomposition

$$r \otimes s = |r - s| \oplus |r - s| + 1 \oplus \ldots \oplus r + s$$

As we saw above, for a Chern-Simons theory $SU(2)_k$, the spin s of the representation must obey $s \leq k/2$. This means that we can't have any representations appearing on

[57] You can find all the details in the yellow "*Conformal Field Theory*" book by Di Francesco, Mathieu and Sénéchal.

the right-hand side which are greater than $k/2$. You might think that we simply delete all representations in the tensor product that are too large. However, it turns out that the fusion rules are more subtle than that; sometimes we need to delete some of the representations that appear to be allowed. The correct fusion rule is

$$r \star s = |r - s| \oplus \ldots \oplus \min(k - r - s, r + s) \tag{5.61}$$

As an example, let's look at $SU(2)_2$. From (5.60), we see that there are just three possible representations, with spin $j = 0, 1/2$ and 1. We'll label these representations by their dimension, $\mathbf{1}, \mathbf{2}$ and $\mathbf{3}$. The fusion rules (5.61) in this case are

$$\mathbf{2} \star \mathbf{2} = \mathbf{1} \oplus \mathbf{3} \quad , \quad \mathbf{2} \star \mathbf{3} = \mathbf{2} \quad , \quad \mathbf{3} \star \mathbf{3} = \mathbf{1} \tag{5.62}$$

Note that the first two of these follow from standard Clebsh-Gordan coefficients, throwing out any spins greater than 1. However, the final product does not have the representation $\mathbf{3}$ on the right-hand side which one might expect. We've seen the fusion rules (5.62) before: they are identical to the fusion of Ising anyons (4.24) with the identification

$$\mathbf{2} \rightarrow \sigma \quad \text{and} \quad \mathbf{3} \rightarrow \psi$$

Recall that these describe the anyonic excitations of the Moore-Read state. Similarly, one can check how many singlets you can form from n spin-1/2 with the requirement that no group has spin greater than 1. The answer, for n even, is $2^{n/2-1}$. We recognise this as the dimension of the Hilbert space of n Ising anyons. This leads us to suspect that the $SU(2)_2$ Chern-Simons theory plays some role in the description of the $\nu = 5/2$ quantum Hall state. We'll look at this in more detail shortly.

Fusion Rules for $SU(N)_k$

For $SU(N)_k$, the fusion rules are simplest to explain using Young diagrams. However, like many aspects of Young diagrams, if you don't explain where the rules come from then they appear totally mysterious and arbitrary, like a weird cross between sudoku and tetris. Here we're not going to explain. We're just going to have to put up with the mystery[58].

[58] A simple mathematica package to compute fusion rules, written by Carl Turner, can be found at http://blog.suchideas.com/2016/03/computing-wzw-fusion-rules-in-mathematica/

We start by writing down the usual tensor product of representations. For each representation on the right-hand side, we draw the corresponding Young diagram and define

$$t = l_1 - k - 1$$

where, as before, l_1 is the length of the first row. Now we do one of three things, depending on the value of t.

- $t < 0$: Keep this diagram.
- $t = 0$: Throw this diagram away.
- $t > 0$: Play. First, we remove a boundary strip of t boxes, starting from the end of the first row and moving downwards and left. Next, we add a boundary strip of t boxes, starting at the bottom of the first column and moving up and right.

  If the resulting Young diagram does not correspond to a representation of $SU(N)$, we throw it away. Otherwise, we repeat until the resulting diagram has $t \leq 0$. If $t = 0$, we again throw it away. However, if $t < 0$ then we keep it on the right-hand side, but with a sign given by

  $$(-1)^{r_- + r_+ + 1} \tag{5.63}$$

  where r_- is the number of columns from which boxes were removed, while r_+ is the number of columns which had boxes added.

An Example: $SU(2)_2$ **Again**

This probably sounds a little baffling. Let's first see how these rules reproduce what we saw for $SU(2)$. We'll consider $SU(2)_2$ which, as we saw, has representations **1**, **2** and **3**. In terms of Young diagrams, these are 1, $\square$ and $\square\square$. Let's look at some tensor products. The first is

$$\mathbf{2} \otimes \mathbf{2} = 1 \oplus \mathbf{3} \quad \Rightarrow \quad \square \otimes \square = 1 \oplus \square\square$$

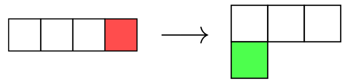
Both boxes on the right-hand side have $t < 0$ so remain. In this case, the fusion rules are the same as the tensor product: $\mathbf{2} \star \mathbf{2} = 1 \oplus \mathbf{3}$. The next tensor product is

$$\mathbf{2} \otimes \mathbf{3} = \mathbf{2} \oplus \mathbf{4} \quad \Rightarrow \quad \square \otimes \square\square = \square \oplus \square\square\square$$

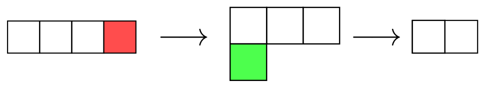
In this case, the final diagram is not an allowed representation of $SU(2)_2$. It has $t = 3 - 2 - 1 = 0$ so we simply discard this diagram. We're left with the fusion rule $\mathbf{2} \star \mathbf{3} = \mathbf{2}$. The final tensor product is

$$\mathbf{3} \otimes \mathbf{3} = 1 \oplus \mathbf{3} \oplus \mathbf{5} \quad \Rightarrow \quad \square\square \otimes \square\square = 1 \oplus \square\square \oplus \square\square\square\square \tag{5.64}$$

The first two diagrams have $t < 0$ and we leave them be. But the third has $l_1 = 4$ and so $t = 1$. This means we can play. We remove a single box from the far right-hand end and replace it below the first box on the left:

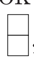
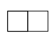
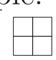
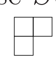


But a column of 2 boxes can be removed in $SU(2)$ Young diagrams. So the full result is

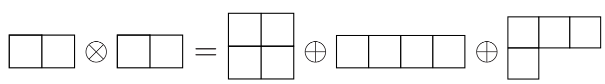


This is another $\mathbf{3}$ representation. But we should worry about the sign. The red box covers a single column, so $r_- = 1$, while the green box also covers a single column so $r_+ = 1$. This means that this diagram comes with a sign -1. This cancels off the  that appeared on the right-hand side of (5.64). This final result is $\mathbf{3} \star \mathbf{3} = 1$. In this way, we see that our rules for manipulating Young diagram reproduce the $SU(2)_2$ fusion rules for Ising anyons (5.62) that we introduced previously.

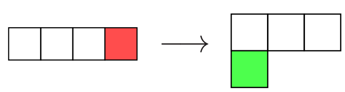
Another Example: $SU(3)_2$

Let's now look another example. We choose $SU(3)_2$. The allowed representations are $\mathbf{3} =$ , $\overline{\mathbf{3}} =$ , $\mathbf{6} =$ , $\overline{\mathbf{6}} =$  and $\mathbf{8} =$ . Let's look at a simple example. The tensor product of two symmetric representations is

$$\mathbf{6} \otimes \mathbf{6} = \overline{\mathbf{6}} \oplus \mathbf{15} \oplus \overline{\mathbf{15}} \quad \Rightarrow$$

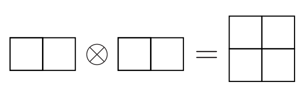


The first of these diagrams has $t < 0$. We keep it. The last of these diagrams has $t = 0$. We discard it. More interesting is the middle diagram which has $t = 1$. This we play with. We have the same manipulations that we saw in the $SU(2)_2$ case above,



However, this time the two boxes in a single column don't cancel because we're dealing with $SU(3)$ rather than $SU(2)$. In fact, as we have seen, this diagram has $t = 0$. We should just discard it. The upshot is that the fusion rules are simply

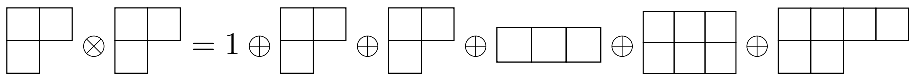
$$\mathbf{6} \star \mathbf{6} = \overline{\mathbf{6}} \quad \Rightarrow$$



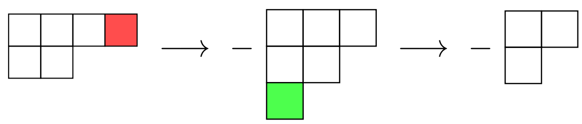
Let's look at another example. The tensor product for two adjoints is

$$\mathbf{8} \otimes \mathbf{8} = \overline{\mathbf{1}} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{10} \oplus \overline{\mathbf{10}} \oplus \mathbf{27}$$

which, in diagrams, reads

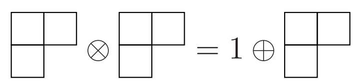


The first three diagrams we keep. The **10** and $\overline{\mathbf{10}}$ diagrams have $t = 0$ and we discard. This leaves us only with the final **27** diagram. This we play with. Using the rules above, we have



where we've now included the minus sign (5.63) in this expression, and the final step comes from removing the column of three boxes. The net result is that the **27** diagram cancels one of the **8** diagrams in the tensor product. We're left with the $SU(3)_2$ fusion rule

$$\mathbf{8} \star \mathbf{8} = \mathbf{1} \oplus \mathbf{8} \quad \Rightarrow$$



We recognise this as the fusion rule for Fibonacci anyons (4.22). This means that the adjoint Wilson lines in $SU(3)_2$ Chern-Simons theory acts like Fibonacci anyons.

Braiding Revisited

We've seen above that Wilson lines in non-Abelian Chern-Simons theories provide an arena to describe non-Abelian anyons. There is a finite dimensional Hilbert space arising from a process of fusion. The next step is obviously to understand braiding in this framework. The adiabatic motion of one Wilson line around another will give rise to a unitary operator on the Hilbert space. How can we calculate this?

There is a long and beautiful story behind this which we will not describe here. The essence of this story is that the action of braiding on the Hilbert space can be translated into the computation of Wilson lines on $\mathbf{S}^3$,

$$\langle W_R \rangle = \int \mathcal{D}a \; e^{iS_{CS}} W_R[a]$$

where R describes the representation of the Wilson line which now traces out some closed, non-intersecting path γ in $\mathbf{S}^3$. In general, such a path describes a tangled path known as a *knot*. Witten famously showed that the expectation value of the Wilson line provides an invariant to distinguish different knots. For $G = SU(2)$, with R the fundamental representation, this invariant is the *Jones Polynomial*.

5.4.5 Effective Theories of Non-Abelian Quantum Hall States

It is clear that non-Abelian Chern-Simons theories give rise to non-Abelian anyons. Indeed, as we mentioned above, for $SU(2)_2$, the structure of anyons that arise is identical to the Ising anyons that describe the Moore-Read states. It's therefore very natural to think they provide effective field theories for the non-Abelian quantum Hall states. And this turns out to be correct. One can argue[59] that the $SU(2)_2$ theory effectively captures the braiding of anyons in the bosonic Moore-Read state at $\nu = 1$.

However, the full description is somewhat involved. One very basic problem is as follows: to construct the full low-energy theory one should identify the electromagnetic current which couples to the background field A_μ. And here gauge invariance works against us. The kind of trick that we used in the Abelian theory is not available for the non-Abelian theory since $\epsilon^{\mu\nu\rho} f_{\nu\rho}$ is not gauge invariant, while $\epsilon^{\mu\nu\rho}\mathrm{tr} f_{\nu\rho} = 0$.

The way to proceed is to look at $U(N) = U(1) \times SU(N)/\mathbf{Z}_N$ Chern-Simons theories. The background gauge field can easily couple to the $U(1)$ factor but we then need the $U(1)$ factor to couple to the rest of $SU(N)$ somehow. This is the part which is a little involved: it requires some discrete identifications of the allowed Wilson lines in a way which is compatible with gauge invariance[60].

However, the Chern-Simons theories also provide us with another way to look at quantum Hall states since these theories are intimately connected to $d = 1 + 1$ dimensional conformal field theories. And it will turn out that these conformal field theories also capture many of the interesting aspects of quantum Hall physics. In our final section, we will look at this for some simple examples.

[59]The argument can be found in "*A Chern-Simons effective field theory for the Pfaffian quantum Hall state*" by E. Fradkin, C. Nayak, A. Tsvelik and F. Wilczek, Nucl.Phys. **B516** 3 704 (1998), cond-mat/9711087.

[60]To my knowledge, this was first explained in Appendix C of the paper by Nati Seiberg and Edward Witten, "*Gapped Boundary Phases of Topological Insulators at Weak Coupling*", arXiv:1602.0425.