Topological Quantum: Lecture Notes

S. Simon

Michaelmas 2016

These are my typed lecture notes. I'm typing as we go so please forgive all typos and unclear parts etc!

General comment. I will not be as precise as mathematicians usually want. Hopefully I will footnote place where important things are swept under the rug.

I'm not sure if these notes will become a book or not. In some ways the notes for a course are not necessarily the right outline for making a good book.

Contents

Chapter 1

Introduction and History of Topology and Kelvin

Long story about Lord Kelvin and Peter Tait. Kelvin thought atoms were knotted vortices in the aether – like smoke rings. Blah blah blah. Tait dies sad :-(

Chapter 2

Kauffman Knot Invariant and Relation to Physics

2.1 The idea of a knot invariant

Topological equivalence. We say two knots are topologically equivalent if they can be deformed smoothly into each other without cutting[1].

While it appears simple to determine whether two simple knots are topologically equivalent and when they are not, for more complicated knots, it becomes *extremely difficult*. (Maybe give an example?)

A **Knot Invariant** is a mapping from a knot (or a picture of a knot) to an output via a set of rules. The rules must be cooked up in such a way that two topologically equivalent knots give the same output.

(See Fig. 2.1)

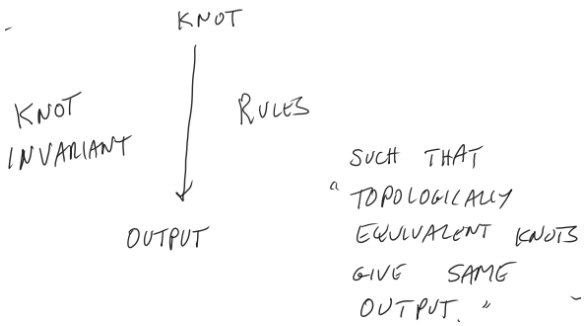


Figure 2.1:

To demonstrate how knot invariants work, we will use the example of the Kauffman invariant[2]. The Kauffman invariant was actually invented by V. Jones who won the Fields medal for his work on knot theory. Kauffman explained his work in very simple terms. Kauffman also wrote a very nice book "Knots and Physics" which I recommend.

[1] A few pieces of fine print here. (1) I am not precise about knot versus link. Strictly speaking a knot is a single strand, and a link is more generally made of multiple strands. Physicists call them all knots. Sorry. (2) When I say "topologically equivalent" heree I mean the concept of regular isotopy. This asks the question of whether there is a continuous smooth family of curves from the initial knot to the final knot – however to be more precise, as we will see below, we should think of the curves as being thickened to ribbons

[2] Be warned, there are multiple things that are called the Kauffman invariant. The one we want is the "the bracket polynomial" multiplied by d

To define the **Kauffman Invariant**, we first choose a number, A. For now, leave it just a variable, although later we may give it a value. There are two rules to the Kauffman invariant. First, a simple loop of string (with nothing going through it) can be removed from the diagram and replaced with the coefficient

$$d = -A^2 - A^{-2}$$

The second rule replaces a diagram with a crossing of strings by a sum of two diagrams where these strings don't cross — where the two possible uncrossings are weighted by A and A^{-1} respectively. The Kauffman rules are shown in Fig. 2.2.

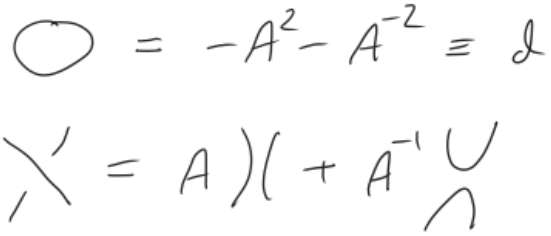


Figure 2.2: Rules for evaluating the Kauffman Invariant

To given an example of how these rules work we show evaluation of the Kauffman invariant for a simple knot in figure 2.3. The output of the calculation is that the Kauffman invariant of this knot comes out to be d. This results is expected since we know that the orginal knot (in the upper left of the figure) is just a simple loop (the so-called "unknot") and the Kauffman rules tell us that a loop gets a value d. We could have folded over this knot many many times[3] and still that outcome of the Kauffman evaluation would be d.

The idea of a knot invariant seems like a great tool for distinguishing knots from each other. If you have two complicated knots and you do not know if they are topologically equivalent, you just plug them into the Kauffman machinery and if they don't give the same output then you know immediately that they cannot be deformed into each other without cutting[4]. However, a bit of thought indicates that things still get rapidly difficult for complicated knots. In the example of fig. 2.3 we have two crossings, and we ended up with 4 diagrams. If we had a knot with N crossings we would have gotten 2^N diagrams, which can be huge! While it is very easy to draw a knot with 100 crossings, even the world's largest computer would not be able to evaluate the Kaufffnman invariant of this knot! So one might then think that this Kauffman invariant is actually not so useful for complicated knots. We will return to this issue later in section 2.4.

2.2 Relation to Physics

There is a fascinating relationship between the Kauffman invariant and quantum physics. For certain types of so-called "topological quantum systems" the amplitudes of space-time processes can be directly calculated via the Kauffman invarient.

We should first comment that most of what we will discuss in this book corresponds to 2 dimensional systems plus 1 dimension of time. There are topological systems in 3+1 dimension as well, but more is known about 2+1 D and we will focus on that at least for now.

Figure 2.4 shows a particular space-time process of particle world lines. At the bottom of the figure is shown the shaded 2 dimensional system (a disk). At some early time there is a pair creation event – a particle antiparticle appear from the vacuum, then another pair creation event; then one particle walks around another, and the pairs come back together to reannihilate.

[3] To a mathematician the Kauffman invariant is an invariant of regular isotopy.

[4] The converse is not true. If two knots give the same output, they are not necessarily topologically equivalent.

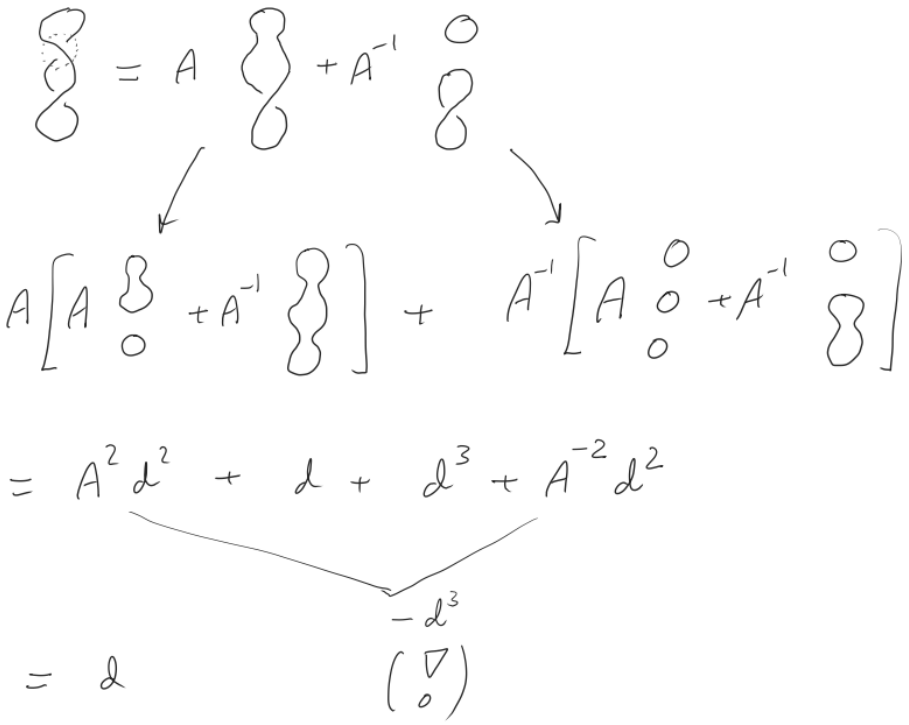


Figure 2.3: Example of evaluation of the Kauffman invariant for the simple twisted loop in the upper left. The light dotted loop is meant to draw attention to where we apply the Kauffman crossing rule first to get the two diagrams on the right hand side. After applying the Kauffman rules again, we have removed all crossings, we are left only with simple loops, which each get the value d. In the penultimate line we have used the definition of d to replace $A^2 + A^{-2} = d$. The fact that we get d in the end of the calculation is expected since we know that the orginal knot is just a simple loop (the so-called "unknot") and the Kauffman rules tell us that a loop gets a value d

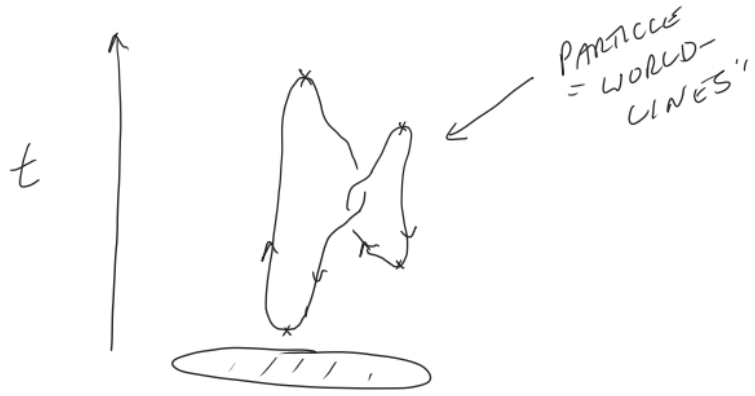


Figure 2.4: A space-time proces showing world lines of particles for a 2+1 dimensional system (shown as the shaded disk at the bottom).

In a topological theory, the quantum amplitude for this depends on the topology of the world lines, and not on the detailed geometry. In other words, as long as the topology of the world lines looks like two linked rings, it will have the same quantum amplitude as that shown in figure 2.4. It should surprise us that such things exist in the real world, as we are used to the idea that amplitudes depend on details of things, like details of the Hamiltonian, how fast the particles move, and how close they come together. But in a topological theory, none of these things matter. What matters is the topoogy of the space-time paths.

What should be obvious here is that the quantum amplitude of a process is therefore a knot invariant. It is a mapping from a knot (made by the world lines) to an output (the amplitude) which depends only on the topology of the knot. This connection between quantum systems and knot invariants was made famously by Ed. Witten, one of the world's leading string theorists. He won the Fields medal along with Vaughn Jones for this work.

Such topological theories were first considered as an abstract possibility, mainly coming from researchers in quantum gravity (we will say a bit more about the relation to quantum gravity later)/ However, now seveal systems are known in condensed matter which actually behave like this. While not all topological theories are related to the Kauffman invariant, many of them are (there are other knot invariants that occur in physical systems as well — including the famous HOMFLY invariant). Let us make a brief table of some of the physical systems that are believed to be related to nontrivial knot invariants:

Table of some interesting topological systems related to knot invariants

1. $SU(2)_2$ **Kauffman class**. For these, the Kauffman invariant gives the quantum amplitude of a process plugging in a value $A = ie^{-i\pi/(2(2+2))} = i^{3/4}$ This includes
   - $\nu = 5/2$ Fractional Quantum Hall Effect (2D electrons at low temperature in high magnetic field). We will say more about FQHE later.
   - 2D Films of exotic superconductors, particularly Sr_2RuO_4
   - 2D Films of ^{3}HeA superfluid[5]
   - A host of "engineered" structures that are designed to have these interesting topological properties. Typically these have a combination of spin-orbit coupling, superconductivity, and magnetism of some sort. Recent experiments have been quite promising
2. $SU(2)_3$ **Kauffman class**. For this, the Kauffman invariant gives the quantum amplitude of a process plugging in a value $A = ie^{-i\pi/(2(2+3))} = i^{4/5}$. The only physical system known in this class is the $\nu = 12/5$ Fractional quantum hall effect.
3. $SU(2)_4$ **Kauffman class**. For this, the Kauffman invariant gives the quantum amplitude of a process plugging in a value $A = ie^{-i\pi/(2(2+4))} = i^{5/6}$. It is possible that $\nu = 2+2/3$ Fractional quantum hall effect is in this class.
4. $SU(3)_2$ **Class**. This corresponds the the HOMFLY knot invariant rather than the Kauffman invariant. It is possible that the unpolarized $\nu = 2 + 4/7$ Fractional quantum hall effect is in this class.

In addition there are a host of complicated systems that could in principle be engineered but are much too hard for current technology to contemplate. There are also other quantum hall states that are also topological, but their corresponding knot invariants are fairly trivial, as we will later see.

2.2.1 Twist and Spin-Statistics

Before moving on, we do a bit of more careful examination of the kauffman invariant. To this end, let us examine a small loop in a piece of string and try to evaluate its kauffman invariant. See figure 2.5.

[5]Two nobel prizes have been given for work on Helium-3 superfluidity.

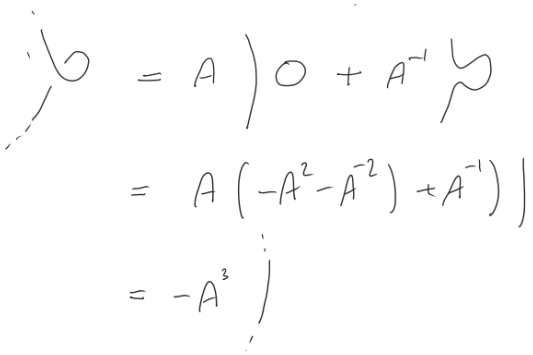


Figure 2.5: Evaluation of a loop. Off the diagram on the left we mean that the string will be connected up with itself, but we are not concerned with any part of the knot except this little piece.

We see from the calculation, that the little loop in the string has value of $-A^3$. But wait! I claimed earlier that any two knots that can be deformed into each other without cutting should have the same Kauffman invariant, but they don't!

Of course I didn't lie to you! The unlooped string on the left and the looped string on the right are not topologically equivalent[6]. To see this we should think of the string as not being infinitely thin, but instead having some width, like a garden hose, or a "ribbon". If we imagine straightening a thick string (not an infinitely thin string) we realize that pulling it straight gives a twisted string — anyone who has tried to straighten a garden hose will realize this! See fig 2.6

Figure 2.6: Pulling straight a small loop

So the loop string is equivalent to a string with a self-twist, and this is then related to a straight string by the factor of $-A^3$. In fact, this is a result we should expect in quantum theory. The string with a self-twist represtents a particle that stays in place but rotates around an axis. In quantum theory, if a particle has a spin, it should accumulate a phase when it does a 2 π rotation, and indeed this factor of $-A^3$ is precisely such a phase in any well defined quantum theory.

In fact, figure 2.6 is a very slick proof of the famous spin statistics theorem. In the left picture with the loop, we have two identical particles that change places. When we pull this straight, we have a single particle that rotates around its own axis. In quantum theory, the phases accumulated by these two processes must be identical. As we will see below, in 2+1D this phase can be arbitrary (not just +1, or -1), but the exchange phase (statistical phase) and the twist phase (the spin phase) must be the same[7].

As a side comment, one can easily construct a knot invariant that treats the looped stirng on the left of Fig.2.5 as being the same as the straight piece of string. One just calculates the kauffman invariant and removes a factor of $-A^3$ for each self twist that occurs. This gives the famed Jones Polynomial knot invariant.

[6] In mathematics we say they are ambient isotopic but not regular isotopic!

[7] As we will see later, there may be multiple exchange phases for two particles, although this does not effect the equivalence of diagrams stated here.

2.3 Bras and Kets

For many topological theories (the so-called non-abelian theories) the physical systems have an interesting, and very unusual property. Imagine we start in a the ground state (or vacuum) state of some systems and create two particle-hole pairs, and if we tell you everything that you can locally measure about these particles (their positions, if they have spin, their spin, etc etc). For most gapped systems (insulators, superconductors, charge density waves) once you know all of the locally measurable quantities, you know the full wavefunction of the system. But this is not true for topological systems. As an example, see figure 2.7.

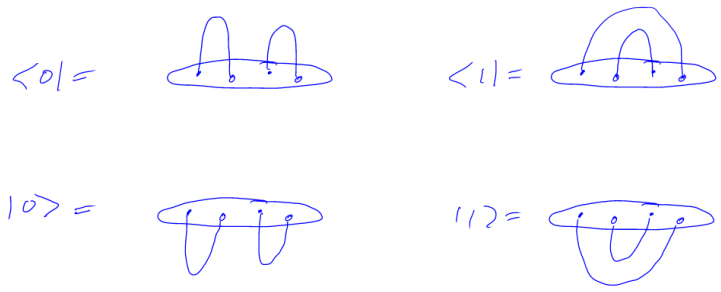


Figure 2.7: Two linearly independent quantum states that look identical locally but have different space-time history.

To demonstrate that these two different space-time histories are linearly independent quantum states, we simply take inner products as shown in Fig.2.8, and we see that $|0\rangle$ and $|1\rangle$ must be linearly independent, at least for $|d| \neq 1$. (We also see that the kets here are not properly normalized, we should multiply each bra and ket by $1/\sqrt{d}$ in order that we have normalized states.)

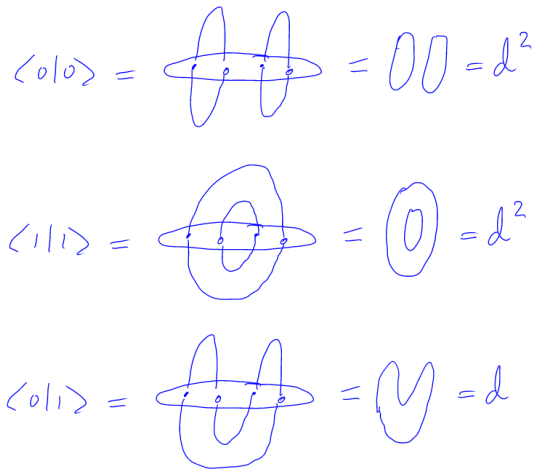


Figure 2.8: Showing that the kets $|0\rangle$ and $|1\rangle$ are linearly indendent. For $|d| \neq 1$ the inner products show they must be linearly independent quantities.

We can even do something more complicated, like insert a braid between the bra and the ket. See Figure 2.9. The braid does a unitary operation on the two dimensional vector space spanned by $|0\rangle$ and $|1\rangle$.

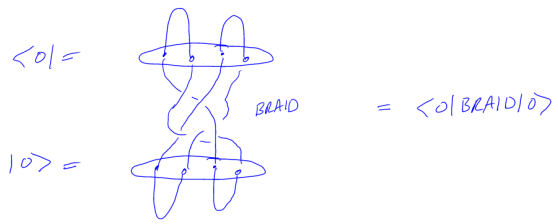


Figure 2.9: Inserting a braid between the bra and the ket. The braid does a unitary operation on the two dimensional vector space spanned by $|0\rangle$ and $|1\rangle$

We emphasize here that these maniplations are not just graphical tricks, but are quite physical. We can think of the $|0\rangle$ and $|1\rangle$ states as being particular operators that produce particle-hole pairs from the vacuum, and the inner product drawn by this graphical procedure is precisely the inner product of these two resulting states. This can be reinterpreted as starting from the vacuum, time evolving with the operator that gives $|0\rangle$ then time evolving with the inverse of the operator that produces $|1\rangle$ to return us to the vacuum.

2.4 Quantum Computation with Knots

Why do we care so much about topological systems and knot invariants? Perhaps we have a hint from the fact that we wrote states above as 0 and 1 – they certainly look like qubits![8] Indeed, quantum computers are really nothing more than bits that you can do unitary operations on.

It turns out that many topological quantum systems can *compute* quanitites efficiently that classical computers cannot.

To prove this, suppose you wanted to calculate the Kauffman invariant of a very complicated knot, say with 100 crossings. As mentioned above, a classical computer would have to evaluate 2^{100} diagrams, which is so enormous, that it could never be done. However, suppose you have a topological system of Kauffman type in your laboratory. You could actually arrange to physically *measure* the Kauffman invariant[9], The way we do this is to start with a system in the vacuum state, arrange to "pull" particle-hole (particle-antiparticle) pairs out of the vacuum, then drag the particles around in order to form the desired knot, and bring them back together to reannihilate. See fig.2.10. Some of the particles will reannihilate, and others will refuse to go back to the vacuum. The probability that they all reannihilate is (up to a normalization[10] given by the absolute square of the Kauffman invariant of the knot (since amplitudes are the Kauffman invariant, the square is the probabiltiy). Even estimation of the Kauffman invariant is essentially impossible for a classical computer. However, this is an easy task if you happen to have a topological quantum system in your lab![11] Thus the topological quantum system has computational ability beyond that of a classical

[8] One of my favorite quotes is "Any idiot with a two state system thinks they have a quantum computer." The objective here is to show that we are not just any idiot — that quantum computing this way is actually a good idea!

[9] Perhaps the first statements ever made about a quantum computer were made by the russian mathematician Yuri Manin, in 1980. He pointed out that doing any calcualtion about some complicated quantum system with 100 interacting particles is virtually impossible for a classical computer. Say for 100 spins you would have to find the eigenvalues and eigenvectors of a 2^{100} dimensioanl matrix. But if you had the physical systm in your lab, you could just measure its dynamics and answer certain questions. So in that sense the physical quantum system is able to compute certain quantities, i.e., its own equations of motion, that a classical computer cannot. In the following year Feynman starting thinking along the same lines and asked the question of whether one quantum system can compute the dynamics of another quantum system — which starts getting close to the ideas of modern quantum computation.

[10] If we pull a single particle-hole pair from the vacuum and immediately bring them back together, the probability that they reannihilate is 1. However, the spacetime diagram of this is a single loop, and the Kauffman invariant is d. The proper normalization is that each pair pulled from the vacuum and then returned from the vacuum introduces a $1/\sqrt{d}$ factor in front of the Kauffman invariant.

[11] Something about the detailed algorithm.

computer.

Figure 2.10: To evaluate the Kauffman invariant: Pull particle-hole pairs from the vacuum, drag them around to form the knot and reannhilate. The probability that they all reannhilate to the vacuum is related to the Kauffman invariant. Thus by repeating the measurement, you can make an accurate estimate of the Kauffman invariant.

It turns out that the ability to calculate Kauffman invariant is sufficient to be able to do any **quantum computation**. So one can use this so-called **topological quantum computer** to run algoritms such as Shor's famous factoring (i.e., code breaking algorithm). The idea of using topological systems for quantum computation is due to Michael Freedman and Alexei Kitaev[12].

So it turns out that these topological systems can do quantum computation. Why is this a good way to do quantum computation?[8]. First we must ask about why quantum computing is hard in the first place. In the conventional picture of a quantum computer, we imagine a bunch of two state systems, say spins, which act as our qubits. Now during our computation, if some noise, say a photon, or a phonon, enters the system and interacts with a qubit, it can cause an error or decoherence, which can then ruin your computation. And while it is possible to protect quantum systems from errors (we will see much later how you do this) it is very hard. Now consider what happens when noise hits a topological quantum computer. In this case, the noise may shake around a particle, as shown in Fig. 2.11. However, as long as the noise does not change the topology of the knot, then no error is introduced. So the topological quantum computer is inherently protected from errors. (of course sufficiently strong noise can change the topology of the knot and still cause errors.)

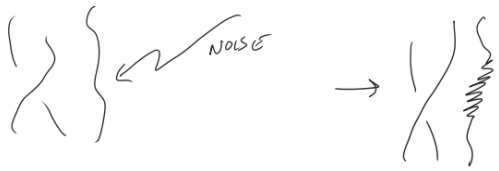


Figure 2.11: The effect of noise on a topological quantum computation. As long as the noise does not change the topology of the knot, then no error is introduced.

2.5 Some quick comments about Fractional Quantum Hall Effect

There will be chapters later about FQHE (if i ever get around to writing them!). But it is worth saying a few words about FQHE as a topological system now.

[12]Freedman is another field's medalist, for his work on the Poincare conjecture in 4d. Alexei Kitaev is one of the most influential scientists alive, a MacArthur winner, Milnor Prize winner, etc. Both smart people.

FQHE occurs in two dimensional electronic systems[13] in high magnetic field at low temperature (typically below 1K). There are many FQHE states which are labeled by their so called filling fraction $\nu = p/q$ with p and q small integers. The filling fraction can be changed in experiment (we will discuss this later). The FQHE state emerges at low temperature and is topological[14]

How do we know that the system is topological. There are not a whole lot of experiments that are easy to do on quantum Hall systems, since they are very low temperature and complicated experiments to do. However, one type of experiment is fairly straightforward — a simple electrical resisitance measurement, as shown in Fig.2.12. In the top of the figure, the so-called longitudinal resistance is measured – where the current runs roughly parallel to the voltage. In this case the measured voltage is zero — like a superconductor. This shows that this state of matter has no dissipation, no friction.

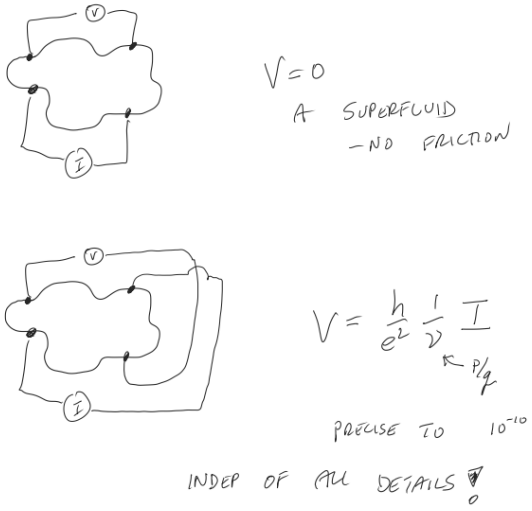


Figure 2.12: Measurement of resistance in FQHE experiment.

The measurement in the lower half of the figure is more interesting. In this case, the Hall voltage is precisely quantized as $V = (h/e^2)(1/\nu)I$ where I is the current, h is Plank's constant,e the electron charge and $\nu = p/q$ is a ratio of small integers. This quantization is extremely precise — to within about a part in 10^{10}. This is like measuring the distance from London to Los Angeles to within a millimeter. Experiments of this sort are used in the metrological definition of the Ohm. What is most surprising is that the measured voltage does not depend on details, such as the shape of the sample, whether there is disorder in the sample, or where you put the voltage leads or how you attach them as long as the current and voltage leads are topologically crossed, as they are in the lower figure, but not in the upper figure. We should emphasize that this is extremely unusual. If you were to measure the resistance of a bar of copper, the voltage would depend entirely on how far apart you put put the leads and the shape of the sample. This extremely unusual independence of all details is a strong hint that we have something robust and topological happening here.

Finally we can ask about what the particles are that we want to braid around each other in the FQHE case. These so-called quasiparticles are like the point-vortices of the FQHE superfluid.

[13]Electronic systems can be made two dimensional in several ways. Most usually electrons are confined in between layers of semiconductors in a so-called heterostructure quantum well. However, one can also use substances like graphene which are only one atom thick and allow electron motion strictly in 2d

[14]A comment in comparing this paradigm to the common paradigm of high energy physics. In high energy there is generally the idea that there is some grand unified theory (GUT) at very high energy scale and it is extremely symmetric, but then when the universe cools to low temperature, symmetry breaks (such as electro-weak symmetry) and we obtain the physics of the world around us. The paradigm is opposite here. The electrons in magnetic field at high temperature have no special symmetry. However, as we cool down to lower temperature, a huge symmetry emerges. The topological theory is symmetric under all diffeomorphisms (smooth distortions) of space and time.

So in fact, Kelvin was almost right! He was thinking about vortices knotting in the dissipationless aether. Here we are thinking about point vortices in the dissipationless FQHE fluid, but we move the vortices around in time to form knots!

Chapter 3

Particle Quantum Statistics

We have been discussing braiding particles around, or exchanging their positions. This is the domain of what we call particle statistics (or quantum statistics). What we mean by this is "what happens to the many particle wavefunction when particles are exchanged in a certain way".

We are familiar with bosons and fermions[1,2]. If we exchange to bosons the wavefunction is unchanged, if we exchange two fermions the wavefunction accumulates a -1 sign. Various arguments have been given as to why these are the only possibilities. The argument usually given in introductory books is as follows

> If you exchange a pair of particles then Exchange them again, you get back where you started. So the square of the exchange operator should be 1, and there two square roots of 1 and +1 and -1. So these are the two possibilities.

In the modern era this argument is considered to be incorrect (or at least not really sufficient). To really understand the possibilities in exchange statistics, we need to think about quantum physics from the Feynman path integral point of view.

Parts of this section might be familiar to many people who know a lot about path integrals. If so, skip down to section 3.2 This is not going to be a course on path integrals, see other refs for that!

3.1 Single Particle Path Integral. Probably you know this!

Consider a space-time trajectory of a non-relativistic particles. We say that we have $\mathbf{x}$ moving in $\mathbb{R}^D$ where D is the dimension of space, so we can write $\mathbf{x}(t)$ where t is time.

Given that we start at position $\mathbf{x}_i$ at the initial time t_i we can define a so-called propagator which gives the amplitude of ending up at position $\mathbf{x}_f$ at the final time t_f. This can be written as

$$\langle \mathbf{x}_f | \hat{U}(t_f, t_i) | \mathbf{x}_i \rangle$$

where $\hat{U}$ is the (unitary) time evolution operator.

The propagator can be used to propagate forward in time some arbitrary wavefunction from t_i to t_f as follows

$$\langle \mathbf{x}_f | \psi(t_f) \rangle = \int d\mathbf{x}_i \, \langle \mathbf{x}_f | \hat{U}(t_f, t_i) | \mathbf{x}_i \rangle \, \langle \mathbf{x}_f | \psi(t_i) \rangle$$

If we are trying to figure out the propagator from some microscopic calculation, there are two very fundamental properties it must obey. First, it must be unitary – meaning no amplitude is lost

[1] Bose cooked up the current picture of Bose statistics in 1924 in the context of photons and communicated it to Einstein who helped him get it published. Einstein realized the same ideas could be applied to non-photon particles as well.

[2] Fermi-Dirac statistics were actually invented by Jordan in 1925. Jordan submitted a paper to a journal, where Max Born was the referee. Born stuck the manuscript in his suitcase and forgot about it for over a year. During that time both Fermi and Dirac published their results. Jordan could have won a Nobel prize (potentially with Born) for his contributions to quantum physics, but he became a serious Nazi and no one really liked him much after that.

along the way (normalized wavefunctions stay normalized). Secondly it must obey compisition... propagating from t_i to t_m and then from t_m to t_f must be the same as propagating from t_i to t_f. We can express the composition law as

$$\langle \mathbf{x}_f | \hat{U}(t_f, t_i) | \mathbf{x}_i \rangle = \int d\mathbf{x}_m \, \langle \mathbf{x}_f | \hat{U}(t_f, t_m) | \mathbf{x}_m \rangle \, \langle \mathbf{x}_m | \hat{U}(t_m, t_i) | \mathbf{x}_i \rangle$$

The integration over $\mathbf{x}_m$ allows the particle to be at any position at the intermediate time (and it must be at *some* position). Another way of seeing this statement is to realize that the integral over $\mathbf{x}_m$ is just insertion of a complete set of states at some intermediate time.

Feynman's genius was to realize that you can subdivide time into infinitessimly small pieces, and you end up doing lots of integrals over all possible intermediate positions. in order to get the final result, you end up summing over all possible values of all possible intermediate positions, or all possible functions $\mathbf{x}(t)$. Feynman's final result is that the propagator can be written as

$$\langle \mathbf{x}_f | \hat{U}(t_f, t_i) | \mathbf{x}_i \rangle = \mathcal{N} \sum_{\substack{\text{paths } \mathbf{x}(t) \text{ from} \\ (\mathbf{x_i}, t_i) \text{ to } (\mathbf{x_f}, t_f)}} e^{iS[\mathbf{x}(t)]/\hbar} \tag{3.1}$$

where $\mathcal{N}$ is some normalization constant. Here $S[\mathbf{x}(t)]$ is the action of the path

$$S = \int_{t_i}^{t_f} dt \, L[\mathbf{x}(\mathbf{t}), \dot{\mathbf{x}}(\mathbf{t}), t]$$

with L the Lagrangian.

The sum over paths in Eq. 3.1 is often well defined as a limit of dividing the path into disctete time steps and integrating over $\mathbf{x}$ at each time. We often rewrite this sum over paths figuratively as a so-called path integral

$$\langle \mathbf{x}_f | \hat{U}(t_f, t_i) | \mathbf{x}_i \rangle = \mathcal{N} \int_{(\mathbf{x_i}, t_i)}^{(\mathbf{x_f}, t_f)} \mathcal{D}\mathbf{x}(t) \; e^{iS[\mathbf{x}(t)]/\hbar} \tag{3.2}$$

Analogous to when we evaluate regular integrals of things that look like $\int dx \, e^{iS[x]/\hbar}$, we can approximate the value of this integral in the small $\hbar$, or classical, limit by saddle point approximation. We do this by looking for a minium of S with respect to its argument — this is where the argument oscillates least, and it becomes the term which dominates the result of the integral. Similarly, with the path integral, the piece that dominates in the small $\hbar$ limit is the piece where $S[\mathbf{x}(\mathbf{t})]$ is extremized — i.e., the classical principle of least action!

3.2 Two Identical Particles

For *identical* particles there is no meaning to saying that particle one is at position $\mathbf{x}_1$ and particle two is at position $\mathbf{x}_1$. This would be the same as saying that they are the other way around. So instead, we can only say that there are particles at positions $\mathbf{x}_1$ and $\mathbf{x}_2$. It is then useful to simply agree on some convention on which coordinate we will write first –for example, maybe we always write the leftmost particle first[3]. For simplicity, we can assume that $\mathbf{x}_1 \neq \mathbf{x}_2$, i.e., the particles have hard cores.

For these indistinguishable particles, the Hilbert space is then cut in half compared to the care of two distinguishable particles where $|\mathbf{x}_1, \mathbf{x}_2\rangle$ and $|\mathbf{x}_2, \mathbf{x}_1\rangle$ would mean physically different things.

The key realization, is that the space of all paths through the configuration space $\mathcal{C}$ divides up into topologically inequivalent pieces. I.e., certain paths cannot be deformed into other paths by a series of small deformations. To the mathematician we are looking at the group of paths through $\mathcal{C}$, known as the first homotopy group $\Pi_1(\mathcal{C})$ or fundamental group.

[3]This ordering scheme works in 1d. In 2d we would perhaps say, the particle with the smaller x coordinate wins, but in case of a tie, the particle with smaller y coordinate wins. etc.

To construct a path integral, we want to think about all possible paths through this configuration space.

Claim: Fixing endpoints, the space of paths through configuration space breaks into topologically disconnected pieces (i..e, they cannot be deformed into each other by a series of small changes.)

What do these topologically disconnected pieces of our space of paths look like? For example, we might consider the two paths as shown in Fig. 3.1. Here we mean that time runs vertically. It is not possible to continuously deform the path on the left into the path on the right assuming the end points are fixed.

Figure 3.1: Two possible sets of paths (paths in configuration space) from the same two starting positions to the same two ending positions. We call the non-exchange path TYPE 1, and the exchange path TYPE -1. Here we mean that time runs vertically. The two sets of paths cannot be continuously deformed into each other assuming the end points are fixed. Note that we may be able to further refine our classification of paths — for example, we may distinguish over and undercrossings, but for now we will only be concerned with exchanges (TYPE -1) and non-exchanges (TYPE 1)

We will call the (left fig) non-exchange path TYPE 1, and the (right fig)exchange path TYPE -1. Here we mean that time runs vertically. The two sets of paths cannot be continuously deformed into each other assuming the end points are fixed. Note that we may be able to further refine our classification of paths — for example, we may distinguish over and undercrossings, but for now we will only be concerned with exchanges (TYPE -1) and non-exchanges (TYPE 1).

Paths can be composed with each other. In other words, we can follow one path, then follow the second. We can write a multiplication table for such composition of paths

TYPE 1	Followed by	TYPE 1	=	TYPE 1
TYPE 1	Followed by	TYPE -1	=	TYPE -1
TYPE -1	Followed by	TYPE 1	=	TYPE -1
TYPE -1	Followed by	TYPE -1	=	TYPE 1

So for example, an exchange path (which switches the two particle) followed by another exchange path (which switches again) results in a net path that does not switch the two particles.

Now let us try to construct a path integral, or sum over all possible paths. It is useful to think about breaking up the sum over paths into seperate sums over the two different classes of paths.

$$\begin{aligned}\langle \mathbf{x}_{1f}\mathbf{x}_{2f}|\hat{U}(t_f,t_i)|\mathbf{x}_{1i}\mathbf{x}_{2i}\rangle &= \mathcal{N}\sum_{\substack{\text{paths}\\ i\to f}} e^{iS[\text{path}]/\hbar} \\ &= \mathcal{N}\left(\sum_{\substack{\text{TYPE 1 paths}\\ i\to f}} e^{iS[\text{path}]/\hbar} + \sum_{\substack{\text{TYPE -1 paths}\\ i\to f}} e^{iS[\text{path}]}\right)\end{aligned}$$

This second line is simply a rewriteing of the first having broken the paths into two different classes.

It turns out however, that it is completely consistent to try something different. Let us instead write

$$\langle \mathbf{x}_{1f}\mathbf{x}_{2f}|\hat{U}(t_f,t_i)|\mathbf{x}_{1i}\mathbf{x}_{2i}\rangle = \mathcal{N}\left(\sum_{\substack{\text{TYPE 1 paths}\\ i\to f}} e^{iS[\text{path}]/\hbar} - \sum_{\substack{\text{TYPE -1 paths}\\ i\to f}} e^{iS[\text{path}]}\right) \quad (3.3)$$

Notice the change of sign for the TYPE -1 paths.

The reason this change is allowed is because it obeys the composition law! To see this (and now using some shorthand), let us check the composition law that we should have. Again, we break the time propagation at some intermediate time

$$\langle \mathbf{x}_{1f}\mathbf{x}_{2f}|\hat{U}(t_f,t_i)|\mathbf{x}_{1i}\mathbf{x}_{2i}\rangle = \int d\mathbf{x}_{1m}d\mathbf{x}_{2m}\,\langle \mathbf{x}_{1f}\mathbf{x}_{2f}|\hat{U}(t_f,t_m)|\mathbf{x}_{1m}\mathbf{x}_{2m}\rangle\,\langle \mathbf{x}_{1m}\mathbf{x}_{2m}|\hat{U}(t_m,t_i)|\mathbf{x}_{1i}\mathbf{x}_{2i}\rangle$$

$$\sim \int d\mathbf{x}_{1m}d\mathbf{x}_{2m}\left(\sum_{\substack{\text{TYPE 1}\\ m\to f}} - \sum_{\substack{\text{TYPE -1}\\ m\to f}}\right)\left(\sum_{\substack{\text{TYPE 1}\\ i\to m}} - \sum_{\substack{\text{TYPE -1}\\ i\to m}}\right) e^{iS[\text{path}]}$$

Now, when we compose together subpaths from $i \to m$ with $m \to f$ to get the overall path, the sub-path types multiply according to our above multiplication table. For the final path, there are two ways to obtain a TYPE 1 path —when either both sub-paths are TYPE 1 or both sub-paths are TYPE -1. In either case, note that the net prefactor of the overall TYPE 1 path is +1. (the two - prefactors of the TYPE -1 multiply and cancel). Similarly, we can get an overall TYPE -1 path. In this case, exactly one of the sub-paths must be of TYPE -1. In which case, the overall sign ends up being -1. Thus, for the full path, we obtain exactly the intended form written in Eq.3.3. I.e., under composition of paths, we preserve the rule that TYPE 1 paths get a +1 sign and TYPE -1 paths get a -1 sign. Thus this is consistent for quantum mechanics, and indeed, this is exactly what happens in the case of fermions.

3.3 Many Identical Particles

To figure out what is consistent in quantum mechanics, we must do two things

(a) Characterize the space of paths through configuration space

(b) Insist on consistency under composition.

Our configuration space for the set of N identical particles in D dimensions can then be written as

$$\mathcal{C} = (\mathbb{R}^{ND} - \Delta)/S_N$$

Here $\mathbb{R}^{ND}$ is a set of N coordinates in D dimensions, Δ is the space of "coincidences" where more than one particle occupy the same position (we are eliminating this possibility for simplicity), here S_N is the group of permutations, and we are "modding" out by this group. We say a bit more about the permutation group in the appendix on group theory, but this modding out by S_N is just a fancy way to say that we specify N coordinates, but we do not order these points (or as described above, we always write the left-most first). In the case of 2 particles above, this reduced the Hilbert space by a factor of 2. More generally this should reduce the Hilbert space by a factor of $N!$. This is the same indistinguishability factor which is familiar from the Gibbs paradox of statistical mechanics.

We would now like to consider all possible paths through this configuration space $\mathcal{C}$. In other words we want to consider how these N different points move in time. We can think of this as a set of coordinates moving through time $\{\mathbf{x}_1(t), \ldots \mathbf{x}_N(t)\}$ but we must be careful that the particles are indistinguishable, so the order in which we write the coordinates doesnt matter.. We can think of this as N directed curves moving in $ND+1$ dimensional space[4]. Since we want to add up all of

[4]The curves are directed because we do not allow them to double back in time, that would represent particle-hole creation or annhilation, which we do not yet consider.

these possible paths in a path integral it is useful to try to better understand what the structure is of this space of paths.

The key realization, is that the space of all paths through the configuration space $\mathcal{C}$ divides up into topologically inequivalent pieces. I.e., certain paths cannot be deformed into other paths by a series of small deformations. To the mathematician we are looking at the group of paths through $\mathcal{C}$, known as the first homotopy group $\Pi_1(\mathcal{C})$ or fundamental group. The reason this is a group (see appendix on definition of a group) is that it comes with a natural operation, or multiplication of elements — which is the composition of paths: follow one path, then follow another path.

3.3.1 Paths in 2+1 D, the Braid Group

An example of a path in configuration space is shown in Fig.3.2. This is known as a braid.

Figure 3.2: A path through configuration space for 3 Particles in 2+1 D is a braid with three strands.

A few notes:

(1) Fixing the endpoints, the braids can be deformed continuously, and so long as we do cut one string through another, it still represents the same topological class.

(2) We cannot allow the strings to double back in time. This would be pair creation or annihilation, which we will consider later, but not now.

The Braid group on N strands is typically notated as B_N. The generators of the braid group on 4 strands are shown in Fig. Any braid can be written as a product of the braid generators

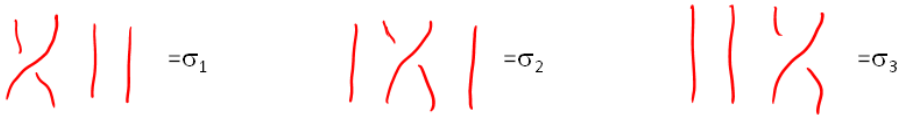


Figure 3.3: The three generating elements of teh braid group on 4 strands. Any braid can be written as a product of the braid generators and their inverses (the inverse of an element looks similir but the crossing is reversed – counter-clockwise instead of clockwise).

and their inverses. An expression representing a braid, such as $\sigma_1\sigma_2\sigma_3^{-1}\sigma_1$ is known as a "braid word." Typically we read the braid word from right to left (do the operation on the right-most first), although sometimes people use the opposite convention!

Note that many different braid words can represent the same braid. An example of this is shown for B_3 in Fig. 3.4. A very useful braid invariant is given by the so-called winding number

$$W\text{=Winding Number} = \#\text{ of overcrossings - }\#\text{ of undercrossings}$$

Where an overcrossig is a σ and an undercrossing is a σ^{-1}. As can be checked in Fig.3.4, the

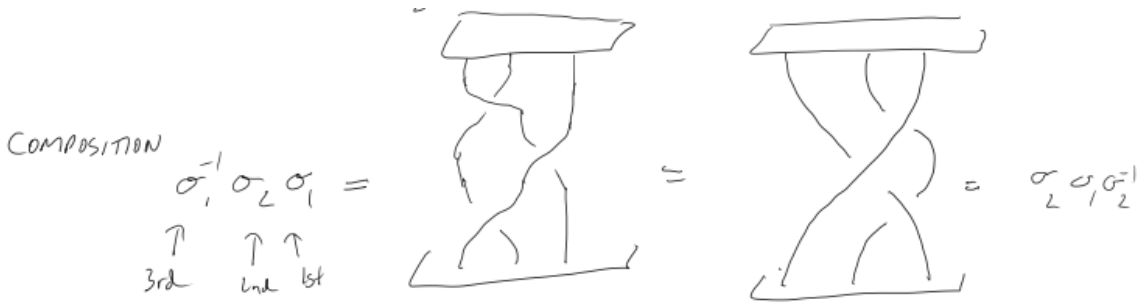


Figure 3.4: Two braid words that represent the same braid.

widing number is independent of the particular way we represent the braid. As long as we do not cut one strand through another or move the endpoints (or double back strands) the braid invariant remains the same.

There are several homework assignments that ask you to use the braid group! Do them and you will be enlightened.

3.3.2 Paths in 3+1 D, the Perumation Group

It would be convenient here to be able to draw pictures in 4 dimensions, but obviously that isn't so easy.

The key point here is that you cannot have any knot in a 1d world-line embedded in a 4d space. If this is not obvious consider the following lower dimensional analogue shown in Fig.3.5.

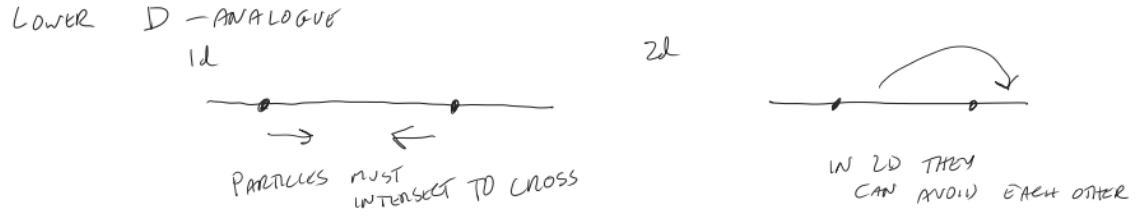


Figure 3.5: In 1d, two points cannot cross through each other without hitting each other. But if we allow the points to move in 2D they can get around each other without touching. This is supposed to show you that 1d world-lines cannot form knots in 4d space.

As shown in the figure, in 1d, two points cannot cross through each other without hitting each other. But if we allow the points to move in 2D they can move around each other without touching each other. Analogously we can consider strings forming knots or braids in 3D space. When we try to push these strings through each other, they bump into each other and get entangled. However, if we allow the strings to move into the 4th dimension, we can move one string a bit off into the 4th dimension so that it can move past the other string, and we see that the strings can get by each other without ever touching each other!

Given that in 3+1D world-lines cannot form knots, the only thing that is important in determining the topological classes of paths is where the strings start and where they end. In other words, we can drraw things that look a bit like braid-diagrams (where the starting plane and finishing plane really represent 3d now!) but now there is no meaning to an over or under-crossing. They are the same! So everything can be unentangled until the diagram looks only like Fig. 3.6. The only thing that is important is who starts where and ends where. This is precisely the permutation group, or symmetric group S_N. Note that in the symmetric group and exchange squared does give the identity.

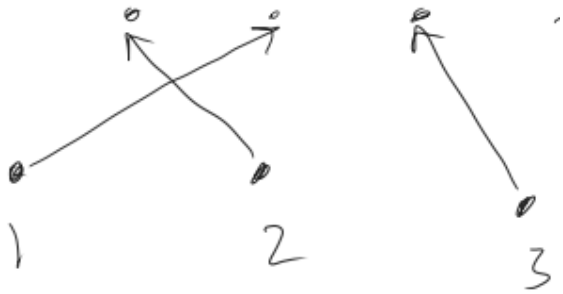


Figure 3.6: Paths in 3+1 D are elements of the permutation group (or symmetric group) S_N.

However, in the braid group this is not so – σ_i^2 is not the identity since it creates a nontrivial braid!

3.3.3 Building a Path Integral

We now return to the issue of building a path integral. We will follow the intuition we gained in the 2-particle case, but now including the information we have discovered about the group of paths through configuration space.

Using the notation $\{\mathbf{x}\}$ to denote all of the N particle coordinates, we now construct the path integral as

$$\langle \{\mathbf{x}\}_f | \hat{U}(t_f, t_i) | \{\mathbf{x}\}_i \rangle = \mathcal{N} \sum_{g \in G} \rho(g) \sum_{\substack{\text{paths} \\ i \to f}} \in g \; e^{iS[\text{path}]/\hbar} \tag{3.4}$$

Here G is the group of paths (the fundamental group) — or the set of classes of topologically different paths. This is the symmetric group S_N for 3+1 dimensions and the braid group B_N for 2+1 dimension . Here we have split the sum over paths into the different classes. We have also introduced a factor of $\rho(g)$ out front where ρ is a *representation* of the group G. (See the appendix on group theory).

To show that this is allowed by the laws of quantum mechanics, we need only check that it obeys the composition law – we should be able to construct all paths from i to f in terms of all paths from i to m and all paths from m to f.

$$\langle \{\mathbf{x}\}_f | \hat{U}(t_f, t_i) | \{\mathbf{x}\}_i \rangle = \int d\{\mathbf{x}\} \; \langle \{\mathbf{x}\}_f | \hat{U}(t_f, t_m) | \{\mathbf{x}\}_m \rangle \; \langle \{\mathbf{x}\}_m | \hat{U}(t_m, t_i) | \{\mathbf{x}\}_i \rangle$$

$$\sim \int d\{\mathbf{x}\} \left(\sum_{g_1 \in G} \rho(g_1) \sum_{\substack{\text{paths} \\ m \to f}} \in g_1 \right) \left(\sum_{g_2 \in G} \rho(g_2) \sum_{\substack{\text{paths} \\ i \to m}} \in g_2 \right) e^{iS[\text{path}]/\hbar}$$

So we have constructed all possible paths from i to f and split them into class g_2 in the region i to m and then class g_1 in the region m to f. When we compose these paths we will get a path of type $g_1 g_2$. The prefactors of the paths $\rho(g_1)$ and $\rho(g_2)$ then multiply and we get $\rho(g_1)\rho(g_2) = \rho(g_1 g_2)$ since ρ is a representation (the preservation of multiplication is the definition of being a representation!). So the prefactor of a given path from i to f is correcly given by $\rho(g)$ where g is the topological class of teh path. In other words, the form shown in Eq. 3.4 is properly preserved under composition, which is what is required in quantum mechanics!

3.4 Abelian Examples

Let us consider the case where the representation ρ of our group G of paths through configuration space is one dimensional – in other words it just a mapping from g to a complex phase.

This case seems to be most applicable in the quantum mechanics we know, because this representation is acting on the wavefunction of our system — and we are quite familiar with the idea of wavefunctions accumulating a complex phase.

3.4.1 3+1 Dimensions

The group of paths through configuration space G is the symmetric group S_N. There are *only two possible* 1-dimensional representations.

Trivial rep: In this case $\rho(g) = 1$ for all g. This corresponds to **bosons**, The path integral is just a simple sum over all possible paths with no factors inserted.

Alternating (or sign) rep: IN this case $\rho(g) = 1$ or -1 depending on whether g represents an even or odd number of exchanges. In this case the sum over all paths gets a positive sign for an even number of exchanges and a negative sign for an odd number. This is obviously **fermions** and is the generalization of the 2-particle example we did above where the exchange was assigned a -1.

3.4.2 2+1 Dimensions

The group of paths through configuration space G is the braid group B_N. We can describe the possible 1-dimensional representations by a single parameter θ. We write the representation

$$\rho(g) = e^{i\theta W(g)}$$

where W is the winding number of the braid g. In otherwords, a clockwise exchange accumulates a phase of e^{iW} whereas a counterclockwise exchange accumulates a phase of e^{-iW}.

For $\theta = 0$ there is no phase, and we simply recover **bosons.**

For $\theta = \pi$ we accumulate a phase of -1 for each exchange no matter the direction of the exchange (since $e^{i\pi} = e^{-i\pi}$). And this is **fermions.**

Any other value of θ is also allowed and this is known as **Anyons**, or fractional statistics. The fact that this is consistent in quantum mechanics was first point out by Leinaas and Myrheim in 1977, and popularized by Wilczek. In 1984,85 Halperin, then Arovas, Schrieffer, and Wilczek showed that anyons really occur in fractional quantum Hall systems.[5]

3.5 Non-Abelian Case

Can we do something more interesting and exotic and use a higher dimensional representation of the group $G = B_N$ of paths in configuration space? Generally in quantum mechanics, higher dimensional representations correpsond to degeneracies, and indeed this is what is necessary.

Suppose we have a system with N particles at a set of positions $\{\mathbf{x}\}$. And even once we fix the positions (as well as the values of any local quantum numbers), suppose there still remains an M-fold degeneracy of the state of the system. We might describe the M states as $|n\rangle$ for $n = 1$ to M. An arbitary wavefunction of the system can then be expressed as

$$|\psi_{\{\mathbf{x}\}}\rangle = \sum_{n=1}^{M} A_n |n; \{\mathbf{x}\}\rangle$$

So given the N positions $\{\mathbf{x}\}$ a general wavefunction should be thought of as a vector in M dimensional space. Now that we have a vector, we can use an M-dimensional representation of the braid group in our path integral! We can use

$$\rho(g) = [U(g)]_{n,n'}$$

[5] A good ref on fractional statisics is the book by Wilczek, "Fractional Statistics and Anyon Superconductivity".

where U is an M by M matrix, and is assumed to be must be a representation of G and it must also be unitary so as to assure that probability is conserved. The unitary matrix will act on the coefficients A_n.

This type of particle is known as a non-Abelian anyon, or nonabelion[6] (since generically matrices don't commute).

In general the Hilbert space dimension M is exponential in the number of particles N. We define a quantity, the **quantum dimension** d

$$M \sim d^N$$

We will see a lot more of this quantity later. It is not coincidence that we used the symbol d previously in the context of Kauffman anyons!

Quantum Computing: Quantum Computing is nothing more than the controlled application of unitary operations to a Hilbert space[7]. Unitary operations is exaclty what one can do by braiding nonabelions around each other!

3.5.1 Parastatistics in 3D

Is it possible to have exotic statistics in 3+1 D? Indeed, there do exist higher dimensional representiongs of the symmetric group. And one can consider particles that obey more complicated statistics even in 3+1 D. However, it turns out that, subject to some "additional constraints", it is essentially not possible to get anything fundamentally new. All we get is bosons and fermions and possibly some internal additional degrees of freedom. The proof of this statement is known as the Doplicher-Roberts theorem (from the 1970s) and runs some 200 pages (so we won't do the whole proof).

However, we should realize that the fine print is important. I mentioned in the previous paragraph that we want to add some "additional constraints" and these are what really limit us to just bosons and fermions. What are these additional constraints. Two things:

(1) We want to be able to pair create and annihilate. This means we are not just considering the braid group, but rather a more complicate structure that allows not just braiding particles around each other, but also creating and annhiliating. This structure is given by category theory, which we will encounter later.

(2) We also want some degree of locality. If we do an experiment on earth, and off on jupiter someone creates a particle-antiparticle pair, we would not want this to effect the result of our experiment at all.

These two restrictions are crucial to reducing the 3+1 D case to only bosons and fermions. We will not go through the details of thi. However, once we see the full structure of particles in 2+1 dimensions, it ends up being fairly clear why the same structure does not work in 3+1 dimensions, except in the case of bosons and fermions.

3.6 Appendix on Groups

It is a shame that the Oxford undergrads do not get any education in group theory. To make up for this failing, I am including a very short exposition of most of what you need to know!

A **group** G is a set of elements $g \in G$ along with an operation that we think of as multiplication. The set is must be closed under this multiplication. So if $g_1, g_2 \in G$ then $g_3 \in G$ where

$$g_3 = g_1 g_2$$

where by writingg $g_1 g_2$ we mean multiply g_1 by g_2. Note: $g_1 g_2$ is not necessarily the same as $g_2 g_1$. If the group is always commutative (ie. $g_1 g_2 = g_2 g_1$ for all $g_1, g_2 \in G$), then we call the group

[6] The idea of nonabelian anyons was explored first in the 1980s by several authors in different contexts. Bais in the context of gauge theories. Froelich in very abstract sense. Witten in the language of topological quantum field theories. Moore and Read in the context of quantum Hall

[7] And initialization and measurement

Abelian[8]. If there are at least some elements in the group where $g_1 g_2 \neq g_2 g_1$ then the group is called **NonAbelian**[9]

A group must always be associative

$$g_1(g_2 g_3) = (g_1 g_2) g_3 = g_1 g_2 g_3$$

Within the group there must exist an **identity** element which is sometimes[10] called e or 0 or 1. The identity element satisifies

$$ge = eg = g$$

for all elements $g \in G$. Each element of the group must also have an inverse which we write as g^{-1} with the property that

$$gg^{-1} = g^{-1}g = e$$

Some Examples

- The group of integers $\mathbb{Z}$ with the operation being addition. The identity element is 0.
- The group $\{1, -1\}$ with the operation being the usual multiplication. This is also called the group $\mathbb{Z}_2$. The identity element is 1. We could have also written this group as $\{0, 1\}$ with the operation being the usual addition modulo 2, where here the identity is 0.
- The group of invertable $n \times n$ complex matrices. We call this group $GL(n, \mathbb{C})$. Here GL stands for "general linear". The identity is the usual identity matrix. By definition all elements of the group are invertable.
- The group of invertable $n \times n$ *real* matrices. We call this group $GL(n, \mathbb{R})$.
- The group of permutations of N elements, which we write as S_N. There are $N!$ elements in the group. Think of the elements of the group as being a one-to-one mapping from the the set of the first N integers into itself. For example, in S_3, one of the elements is

$$g_1 = \begin{cases} 1 & \to & 2 \\ 2 & \to & 1 \\ 3 & \to & 3 \end{cases}$$

Another element is

$$g_2 = \begin{cases} 1 & \to & 2 \\ 2 & \to & 3 \\ 3 & \to & 1 \end{cases}$$

The multiplication operation $g_3 = g_2 g_1$ is meant to mean, do 1 first, then do 2 (you should be careful to make sure your convention of ordering is correct. Choose either convention, but stick to it!) So, if we start with the element 1. Do g_1 the element 1 gets moved to 2. Then when we do g_2 the element 2 gets moved to 3. So in the product $g_2 g_1$ we have 1 getting moved to 3. In the end we have

$$g_3 = g_2 g_1 = \begin{cases} 1 & \to & 3 \\ 2 & \to & 2 \\ 3 & \to & 1 \end{cases}$$

Representations of Groups:

[8] Named after Abel, the Norwegian mathematician who studied such groups in the early 1800s

[9] Apparently named after someone named NonAbel.

[10] It may seem inconvenient that the identity has several names. However, it is sometimes convenient. If we are thinking of the group of integers and the operation of addition, we want to use 0 as the identity. If we are thinking about the group $\{1, -1\}$ with the operation of usual multiplication, then it is convenient to write the identity as 1.

A **representation** is a group homomorphism. Homomorphism means it is a mapping from one group to another which preserves multiplication. We will be concerned with the most common type of representation, which is a homomorphism into the general linear group, ie, the group of matrices. Almost always we will work with complex matrices. Thus our n-dimensional representation is a mapping ρ to n-dimensional complex matrices

$$\rho : G \to GL(n, \mathbb{C})$$

preserving multiplication. I.e.,

$$\rho(g_1)\rho(g_2) = \rho(g_1 g_2)$$

for all $g_1, g_2 \in G$.

Representation theory of groups is a huge subject, but we won't discuss it further here!

Chapter 4

Aharanov-Bohm, Charge-Flux Composites, and Introducing Chern-Simons Theory

4.1 Review of Aharanov-Bohm Effect

Let us consider the two slit experiment shown in Fig. 4.1

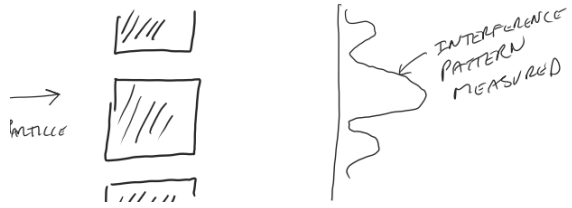


Figure 4.1: The Young two slit experiment

We all know the result of the two slit experiment but let us rewrite the calculation roughly as a path integral. We can write

$$\sum_{\text{paths}} e^{iS/\hbar} = \sum_{\text{paths, slit 1}} e^{iS/\hbar} + \sum_{\text{paths, slit 2}} e^{iS/\hbar}$$
$$\approx\sim \; e^{iL_1/\lambda} + e^{iL_2/\lambda}$$

where L_1 and L_2 are the path lengths through the two respective slits to whichever point is being measured at on the output screen. In other words, we get the usual 2-slit calculation.

Now let us change the experiment to that shown in Fig. 4.2. Here we assume the particle being sent into the interferometer is a charged particle, such as an electron. In this case a magnetic field is added inside the middle box between the two paths. No magnetic field is allowed to leak out of the box, so the particle never experiences magnetic field. Further the magnetic field is kept constant so the particle does not feel a Faraday effect either. The surprising result is that the presence of the magnetic field nonetheless changes the interference pattern obtained on the observation screen! This effect, named the Aharanov-Bohm effect, was predicted by Ehrenberg and Siday in 1949, then re-predicted by Aharanov and Bohm in 1959[1].

[1]Possibly the reason it is named after the later authors is that they realized the importance of the effect, whereas the earlier authors pointed it out, but did not emphasize how strange it is! The first experimental observation of the effect was in 1960 by Chambers, although many more careful experiments have been done since.

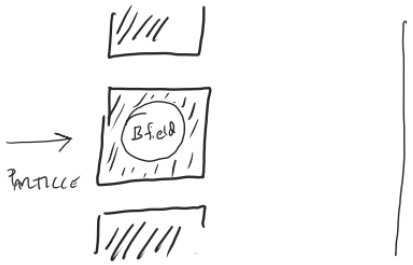


Figure 4.2: Adding a magnetic field into the hidden box in the Young two slit experiment. Here the circular region includes a constant magnetic field. No magnetic field leaks out of the box. Nonetheless, if the particle being sent into the interferometer is charged, the interference pattern is changed compared to the above figure.

So why does this strange effect occur? There are several ways to understand it, but the best for our purpose will be to stay with the idea of path integrals and consider the Lagrangian description of particle motion.

We must recall how a charged particle couples to an electromagnetic field in the Lagrangian description of mechanics. We write the magnetic field and electric field in terms of a vector potential

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

and the electric field will then be

$$\mathbf{E} = -\nabla A_0 - d\mathbf{A}/dt$$

where A_0 is just the electrostatic potential. We can then write the particle Lagrangian as

$$L = \frac{1}{2m}\dot{\mathbf{x}}^2 + q(\mathbf{A}(\mathbf{x}) \cdot \dot{\mathbf{x}} - A_0) \tag{4.1}$$

where q is the particle charge.

It is an easy exercise to check that the Euler-Lagrange equations of motion from this Lagrangian correctly gives motion under the Lorentz for as we should expect in electromagnetism.[2] So adding a magnetic field to the Lagrangian can be rephrased as changing the action as

$$\begin{aligned} S \to \quad & S_0 + q\int dt\ \dot{\mathbf{x}} \cdot \mathbf{A} \\ \to \quad & S_0 + q\int \mathbf{dl} \cdot \mathbf{A} \end{aligned}$$

where S_0 is the action in the absence of the magnetic field.

Returning now to the two slit experiment. The amplitude of the process in the presence of the vector potential can be now rewritten as

$$\sum_{\text{paths, slit 1}} e^{iS_0/\hbar + iq/\hbar \int \mathbf{dl}\cdot\mathbf{A}} + \sum_{\text{paths, slit 2}} e^{iS/\hbar + + iq/\hbar \int \mathbf{dl}\cdot\mathbf{A}}$$

[2]Here are the steps:

$$\begin{aligned} \frac{d}{dt}\frac{\partial L}{\partial \dot{x}_k} &= \frac{\partial L}{\partial x_k} \\ \frac{d}{dt}(m\dot{x}_k + qA_k) = m\ddot{x}_k + q\frac{d}{dt}A_k + q\dot{x}_j\frac{\partial}{\partial x_j}A_k &= q(\dot{x}_j\frac{\partial}{\partial x_k}A_j - \frac{\partial}{\partial x_k}A_0) \\ m\ddot{x}_k &= q(\mathbf{E} + \dot{\mathbf{x}} \times \mathbf{B})_k \end{aligned}$$

Where S_0 is the action of the path in the absence of the vector potential.

The physically important quantity is the difference in accumulated phases between the two paths. This is given by

$$\exp\left[\frac{iq}{\hbar}\int_{slit1}\mathbf{dl}\cdot\mathbf{A}-\frac{iq}{\hbar}\int_{slit2}\mathbf{dl}\cdot\mathbf{A}\right]=\exp\left[\frac{iq}{\hbar}\oint\mathbf{dl}\cdot\mathbf{A}\right] \tag{4.2}$$

where the integral is around a loop that goes forward through slit one and then backwards through slit 2.

Using stokes theorem, we have

$$\frac{iq}{\hbar}\oint\mathbf{dl}\cdot\mathbf{A}=\frac{iq}{\hbar}\int_{inside\,loop}\mathbf{dS}\cdot(\nabla\times\mathbf{A})=\frac{iq}{\hbar}\Phi_{enclosed}$$

where $\Phi_{enclosed}$ is the flux enclosed in the loop. Thus there is a measurable relative phase shift between the two paths given by $\frac{iq}{\hbar}\Phi_{enclosed}$. This results in a shift of the interference pattern observed on the observation screen. Note that although the orginal Lagrangian did not look particularly gauge invariant, the end result (once we integrate around the full path) is indeed gauge independent.

A few notes about this effect:

(1) If Φ is an integer multiple of the elementary flux quantum

$$\phi_0=2\pi\hbar/q$$

, then the phase shift is an integer multipe of 2 π and is hence equivalent to no phase shift.

(2) We would get the same phase shift if we were to move flux around a charge. (Sometimes known as Aharonov-Casher effect)

(3) More generally for particles moving in general space-time one wants to calculate the relativistically invariant quantity

$$\frac{iq}{\hbar}\oint dl_\mu A^\mu$$

4.2 Anyons as Flux-Charge Composites

Good reference is Preskills notes pages 1-25; Also the book on Anyons by Wilczek.

We will now consider a simple model of Abelian anyons — charge-flux composites. Imagine we have a 2 dimensional system with charges q in them, where each charge is bound to an infinitely thin flux tube through the plane, with flux Φ as shown in Fig. 4.3 If we drag one particle around

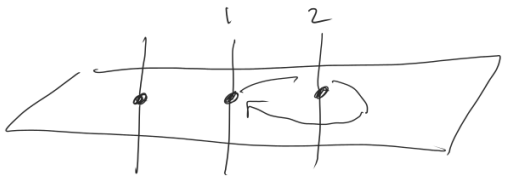


Figure 4.3: Abelian Anyons represented as charges bound to flux tubes through the plane. The charge of each particle is q, the flux of each tube is Φ.

another, we then accumulate a phase due to the Aharnov-Bohm effect. The phase from the charge of particle 1 going around the flux of particle 2 is $e^{iq\Phi/\hbar}$, whereas the phase for dragging the flux of 1 around the charge of 2 is also $e^{iq\Phi/\hbar}$, thus the total phase for dragging 1 around 2 is given by

$$\text{(Phase of flux-charge composite 1 encircling 2)}=e^{2iq\Phi/\hbar}$$

Thus we have (as shown in Fig. 4.4)

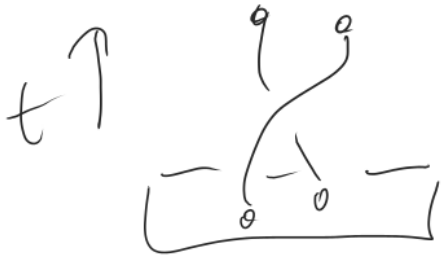


Figure 4.4: An exchange

$$\text{(Phase for exhange of two flux-charge compisites)} = e^{iq\Phi/\hbar}$$

and we correspondingly call these particles θ-anyons, with $\theta = q\Phi/\hbar$. Obviously $\theta = 0$ is bosons, $\theta = \pi$ is fermions, but other values of θ are also allowed.

Spin of an anyon

Let us see if we can determine the spin of these anyons. Spin refers to the rotation operator, so we need to physically rotate the anyon on its axis. To do this we must think about how the flux is tied to the charge — we must have some microscopic description of exactly where the flux is and where the charge is. It is easiest to put them at very slightly different positions[3]. In this case, when we rotate the anyon around its axis we move the charge and flux around each other and we obtain a new phase of

$$e^{1q\Phi/\hbar} = e^{i\theta}$$

This fits very nicely with the spin statistics theorem – the phase obtained by exchanging two identical particles should be the same as the phase obtained by rotating one around its own axis. (See the discussion by Fig. 2.6).

4.2.1 Anti-Anyons

We can introduce the concept of an anti-anyon. This would be a charge flux composite which instead of having charge and flux (q, Φ) instead has $(-q, -\Phi)$. This makes sense because if we pair create an anyon-antianyon pair, before the creation, we have total charge $q = 0$ and total flux $\Phi = 0$, and after the creation the sum of the two charges is still zero, as is the sum of the two fluxes.

If the phase of dragging an anyon clockwise around an anyon is 2θ, then so is the phase of dragging an anti-anyon clockwise around an anti-anyon! (The two minus signs on the two anyons cancel). However, the phase of dragging an anyon clockwise around an antianyon is -2θ.

4.2.2 Fusion of Anyons

We can consider pushing two anyons together to try to form a new particle. A simple example of this is pushing together an anyon with an antianyon. In this case the charge and flux (q, Φ) cancels with the charge and flux $(-q, -\Phi)$ giving the vacuum. This makes some sense — the total charge and flux are locally conserved. (See fig. 4.5).

If we push together two particles both having charge and flux (q, Φ) we will obtain a single particle with charge and flux $(2q, 2\Phi)$. Note that the phase of exchanging two such double particles is now $\theta = 4q\Phi/\hbar$. We sometimes will draw a "fusion diagram" as in Fig. 4.5 to indicate that two θ particles have come together to form a 4θ particle.

[3]We can try to put them at the same position, but it becomes very hard to not get infinities if we do this!

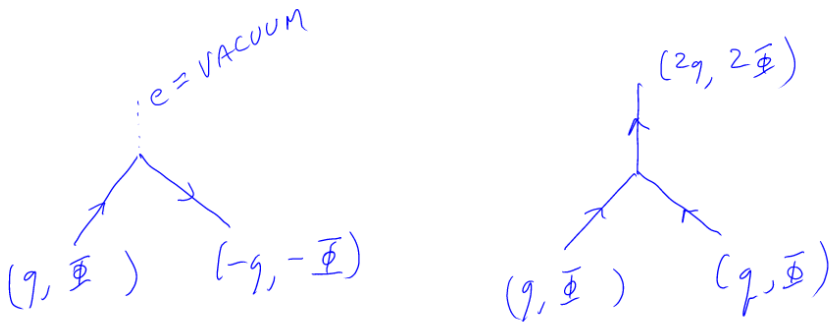


Figure 4.5: Left; Fusing an anyon and an antianyon to get the vacuum ("e") drawn as dotted line. Note that the antianyon moving forward in time is drawn as a downpointing arrow — which looks like an anyon moving backwards in time. Right: Fusing two anyons together to form another anyon of a different type (two $\theta = q\Phi/\hbar$ anyons make a 4θ anyon)

(Add better picture of braiding around the vacuum and braiding around p-h pair giving zero net phase).

The principle of locality is an important one. Consider Fig. 4.6) where we see that braiding around a segment that doubles back in time is equivalent to braiding around a segment pulled straight. This is an example of the locality priniciple. If you wrap around a large region of space, it should not matter if there is a glitch in that region of space where an anyon and antianyon are created briefly. From far away you should not see this. You should only see that the total *anyonic charge* in that region of space is zero.

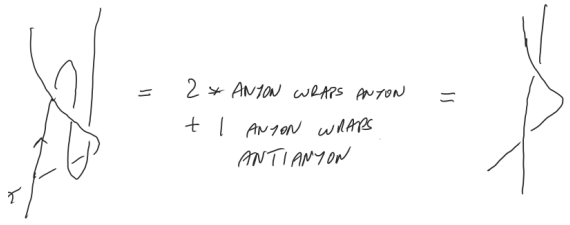


Figure 4.6: Braiding around a world-line that doubles back. The anyon-antianyon braiding rules are constructed so that the double-back can be pulled straight. The calculation scribbled in the middle says "two anyon wraps around anyon + 1 anyon wraps anti-anyon" which give a net phase the same as one anyon wrapping one anyon.

4.3 Anyon Vacuum on a Torus, etc, and Quantum Memory

A rather remarkable feature of topological models is that the ground state somehow "knows" what kind of anyons exist in the model (i.e, those that *could* be created), even when they are not actually present. To see this, consider the ground state of an anyon model on the surface of a torus (a doughnut – or donut if you are from the states).

We can draw the torus as a square will opposite edges identified as shown in Fig. 4.7. The two cycles around the torus are marked as C_1 and C_2.

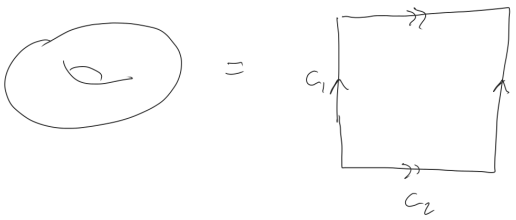


Figure 4.7: Drawing a torus as a rectangle with opposite edges identified. The two noncontractable cycles around the torus can be considered to be the edges of the square, labeled C_1 and C_2 here.

Let us now construct operators that do the following complicated operations:

> T_1 is the operator that creates a particle-hole pair, moves the two in opposite directions around the C_1 cycle of the torus, they meet on the opposite side of the torus and reannihilate.
> T_2 is the operator that creates a particle-hole pair, moves the two in opposite directions around the C_2 cycle of the torus, they meet on the opposite side of the torus and reannihilate.

Both of these operators are unitary because they can be implemented (in principle) with some time-dependent Hamiltonian[4]. However, the two operators do not commute. To see this consider the operator $T_2^{-1}T_1^{-1}T_2T_1$ where I read time from right to left. This can be interpreted as as two particles being created, braiding around each other, and then reannihilating. This is shown in Fig. 4.8.

So what we have now is two operators T_1 and T_2 which do not commute with each other, indeed, we have[5]

$$T_2T_1 = e^{-2i\theta}T_1T_2$$

But both T_1 and T_2 commute with the Hamiltonian (since they start and end with states of exactly the same energy[6]). Whenever you have two operators that don't commute with each other but do commute with the Hamiltonian, it means you have degenerate eigenstates. Let us see how this happens.

Since T_1 is unitary, its eigenvalues have to have unit modulus (or just a complex phase). Consider the space of possible ground states, let us write a ground state eigenstate of T_1 as

$$T_1|\alpha\rangle = e^{i\alpha}|\alpha\rangle$$

Now we can generate a new eigenstate with a different eigenvalue of T_1. Consider the state $T_2|\alpha\rangle$. This must also be in the ground state space since T_2 commutes with the Hamiltonian. But now

$$T_1(T_2|\alpha\rangle) = e^{2i\theta}T_2T_1|\alpha\rangle = e^{2i\theta}e^{i\alpha}(T_2|\alpha\rangle)$$

So we can call this new ground state $|\alpha + 2\theta\rangle = T_2|\alpha\rangle$. So we have now generated a new ground state and we can continue the procedure to generate more!

[4]For example, we could insert charges +Q and -Q near to each other which are strong enough to pull a particle-hole pair out of the vacuum, the -Q trapping the $+(q, \Phi)$ and the +Q trapping the $(-q, -\Phi)$. Then we can drag the Q charges around the handle of the torus, dragging the anyons with them.

[5]At least this relation should be true acting on the ground state space. If some particles are already present, then we have to consider the braiding of the the particles we create with those already present as well, which will be more complicated.

[6]Strictly speaking this means they commute with the Hamilonian within the ground state space.

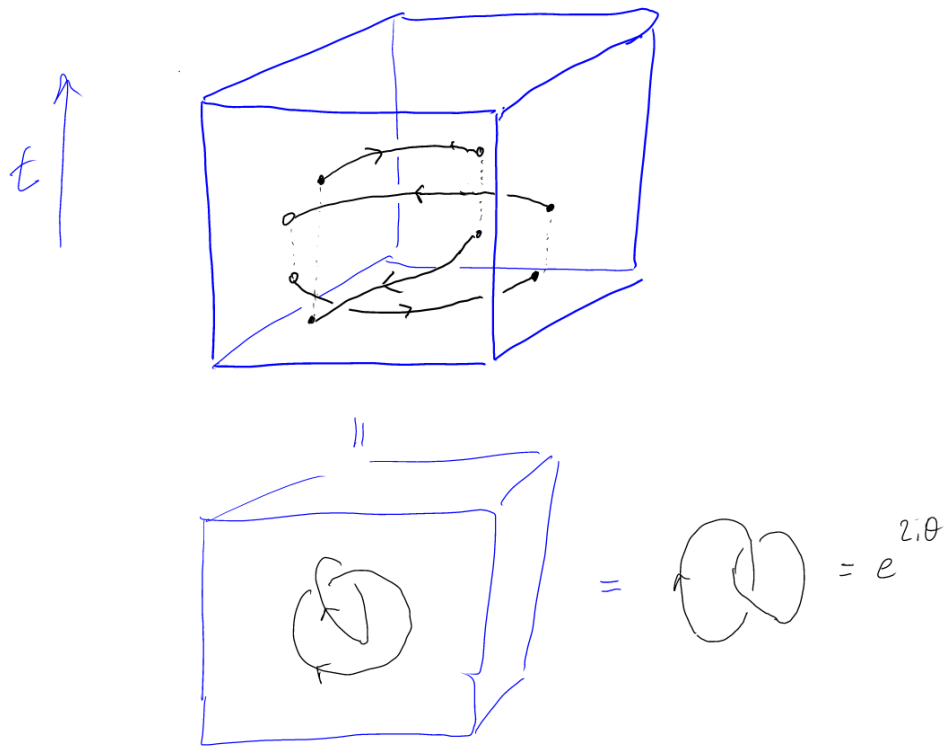


Figure 4.8: The torus is drawn as a horizontal rectangle with opposite ends identified. Time runs vertically. First create a particle hole pair at the center of the rectangle and move them in opposite directions until they meet at the edges of the rectangle to reannhiliate. Note that a particle moving to the right or an antiparticle moving to the left are both drawn as a rightpointed arrow. Similarly next a particle antiparticle pair are made in the center of the torus and moved to the front and back walls (which are the same point) to reannihilate. Then the two processes are reversed to give $T_2^{-1}T_1^{-1}T_2T_1$. This procedure can be reduced to one particle wrapping around another which gives a phase of $e^{-2i\theta}$. Note that in the top figure we do not quite annihilate the particles at the end of the first and second step. This is actually allowed since bringing two particles close looks like they have fused together from far away.

Let us suppose we have a system where the anyons have statistical phase

$$\theta = \pi p/m$$

where p and m are relatively prime integers (i.e., it is an irreducible fraction). Now we have a series of ground states

$$|\alpha\rangle, \qquad |\alpha + 2\pi p/m\rangle, \qquad |\alpha + 4\pi p/m\rangle, \qquad \ldots \qquad , \qquad |\alpha + 2\pi(m-1)/m\rangle$$

When we try to generate yet another state, we get the phase $\alpha + 2\pi = \alpha$ so we are back to the original state. Nonetheless, we now have m independent ground states[7].

Now let us consider the anyons in the system. Since we are considering anyons of statistical angle $\theta = \pi p/m$ we can consider a charge-flux composite with $(q, \Phi) = (\pi p/m, 1)$. Fusion of n of these elementary anyons will have (By this time I'm sick of $\hbar$ and I'm going to set it equal to 1.)

$$\text{Fusion of } n \text{ elementary anyons} = |"n"\rangle = (q = n\pi p/m, \Phi = n) = (n\pi p/m, n)$$

[7]There may be even more degeneracy which is non generic. What we have proven is there must be a degeneracy which is m times some integeer, where one generally expects that integer to be 1 but there could be additional accidental degeneracy.

Now let us consider a cluster of m of these elementary anyons. We then have

$$|"m"\rangle = (\pi p, m)$$

If we braid an arbitrary cluster $|"n"'\rangle = (n\pi p/m, n)$ around one of these $|"m"\rangle = (\pi p, m)$ cluster, we obtain a net phase[8] of $2n\pi p$ which is equivalent to no phase at all! Thus we conclude that the cluster of m elementary anyons is equivalent to the vacuum in the sense that everything braids trivially when it goes all the way around it.

We might be tempted to conclude that there are exactly m different anyon species in the system. Indeed, this conclusion is often true. However, there is a glitch. If both p and m are odd, one obtains a nontrivial sign for exchanging (half braiding) a $|"m"\rangle = (\pi p, m)$ with another $|"m"\rangle = (\pi p, m)$ particle (you get a phase πpm since exchange should give half of the $2\pi pm$ phase for wrapping one particle all the way around the other). This means the $|"m"\rangle$ particle is a fermion. In fact, this case of p and m both odd is a bit of an anomolous case and in some sense is a poorly behaved theory[9].

Neglecting this complicated case with fermions, we are correct to conclude that we have exactly m different species of anyons – and also m different ground states on the torus. This connection will occur in any well behaved topological theory — the number of ground states on the torus will match the number of different types of particles.

Quantum Memory and Higher Genus

The degenerate ground state on the torus can be thought of as a quantum memory. There are m different ground states, so the most general wavefunction we can have is some linear superposition

$$|\Psi\rangle = \sum_{n=0}^{m-1} A_n |\alpha + 2\pi np/m\rangle$$

where the coefficients A_n form an arbitrary (but normalized) vector. We can initialize the system in some particular superposition (i.e, some vector A_n) and we can expect that the system remains in this superposition. The only way that this superposition can change is if a T_1 or T_2 operation is performed (or some combination thereof)— i.e, if a pair of anyons appears from the vacuum moves around the handle of the torus and then reannihilates. Such a process can be extremely unlikey when the energy gap for creating excitations is large. Hence the quantum superposition is "topologically protected".

In fact, one does not even need to have a system on a torus in order to have a degenerate ground state. It is often sufficeint to have an annulus geometry (a disk with a big hole in the middle). In this case, T_1 could correspond to moving an anyon around annulus and T_2 could correspond to moving an anyon from the inside to the outside edge. In this case it is often not precisely true that the ground states are entirely degenerate (since there is a net result of having moved a particle from inside to outside, and therefore one is not necessarily in the precise ground state) but under certain conditions it can be extremely close to degenerate nonetheless.

One can consider mode complicated geometries, such as a torus with multiple handles, or a disk with multiple holes cut in the middle. For a theory of abelian anyons (fractional statistics) the ground state degeneracy for a surface with **genus** g (meaning g handles, or g holes) is m^g. Thus by using high genus one can obtain very very large Hilbert spaces in which to store quantum information.

4.4 Abelian Chern-Simons Theory

Good ref is Wilczek book.

It is useful to see how charge-flux binding occurs in some microscopic field theory description of a physical system. The type of field theory we will study is known as a Chern-Simons field theory[10] and is the paradigm for topological quantum field theories.

[8] $n\pi p/m$ times m plus n times πp.

[9] Later on we will call it "non-modular."

[10] S.S. Chern is one of the most important mathematicians of the 20th century. He died in 2004. Jim Simons was

In the current section we will consider the simplest type of Chern-Simons theory which is the Abelian type (i.e., it generates Abelian anyons). To do this we imagine a gauge field a_α analogous to the vector potential A_α we have discussed already when we were discussing flux above. Here we should realize that a_α is not the real electromagnetic vector potential because it lives only in our 2-dimensional plane. We should think of it instead as some emergent effective quantity for whatever two dimensional system we are working with.

Let us write the Lagrangian of our system

$$L = L_0 + \int d^2x \Delta\mathcal{L}$$

Here we have written L_0 to be the Lagrangian of our particles without considering the coupling to the vector potential. This could be nothing more than the Lagrangian for free particles – although we could jam other things into this part too, such as inter-particle interaction if we like.

The second term is the integral of a Lagrangian density — and this will be the term that is relevant for the flux-binding and the exchange statistics of the particles. The form of the Lagrangian density is

$$\Delta\mathcal{L} = -qj^\alpha a_\alpha + \frac{\mu}{2}\epsilon^{\alpha\beta\gamma}a_\alpha\partial_\beta a_\gamma \tag{4.3}$$

Here q is the particle charge, j^α is the particle current, μ is some coupling constant, and ϵ is the antisymmetric tensor. The indices α, β, γ take values $0, 1, 2$ where 0 indicates the time direction and 1, 2 are the space directions.

The first term here $j^\alpha a_\alpha$ is actually something we have already seen. If we have N particles then the current is

$$j^0(\mathbf{x}) = \sum_{n=1}^{N}\delta(\mathbf{x}-\mathbf{x}_n)$$
$$\vec{j}(\mathbf{x}) = \sum_{n=1}^{N}\dot{\vec{x}}\,\delta(\mathbf{x}-\mathbf{x}_n)$$

The j^0 component, the density, is just a delta function peak at the position of each particle. The 1 and 2 component, $\vec{j}$ is a delta function at the position of each particle with prefactor of the velocity of the particle. Now when $-qj^\alpha a_\alpha$ is integrated over all of space we get

$$\sum_{n=1}^{N} q\left[\mathbf{a}(\mathbf{x}_n)\cdot\mathbf{x}_n - a_0(\mathbf{x}_n)\right]$$

exactly as in Eq. 4.1. So this is nothing more than the regular coupling of a system of charged particles to a vector potential.

The second term in Eq. 4.3 corresponds to the Lagrangian density of the Chern-Simons vector potential itself. (It is sometimes known as the "Chern-Simons Term").

As mentioned above the coupling of the particles to the gauge field is gauge invariant once one integrates the particle motion over some closed path. The Chern-Simons term is gauge invariant on a closed manifold if we can integrate by parts. To see this, make an arbitray gauge transformation

$$a_\mu \to a_\mu + \partial_\mu\chi \tag{4.4}$$

for any function χ. Then integating the Chern-Simons term by parts if necessary all terms can be brought to the form $\epsilon^{\alpha\beta\gamma}\chi\partial_\alpha\partial_\beta a_\gamma$ which vanish by antisymmtry. Note that this this gauge invariance

a prominent mathematician who wrote the key first paper on what became known as Chern-Simons theory in 1974. Simons was the head of the math department at Stonybrook university at the time. In 1982, he decided to change careers and start a hedge fund. His fund, Renaissance Technologies, became one of the most successful hedge funds in the world. Simons' wealth is now estimated at over 16 billion dollars. More recently he has become a prominent philanthropist, and has donated huge amounts of money to physics and mathematics — now being one of the major sources of funds for the best scientists in the US.

holds for any closed manifold, although for a manifold with boundaries, we have to be careful when we integrate by parts as we can get a physically important boundary term. (We will discuss these later, but for now, let us just think about closed space-time manifolds).

To deterimne what the Chern-Simons does we need to look at the Euler-Lagrange equations of motion. We have

$$\frac{\partial \mathcal{L}}{\partial a_\alpha} = \partial_\beta \left(\frac{\partial \mathcal{L}}{\partial(\partial_\beta a_\alpha)} \right)$$

which generates the equations of motion[11]

$$q j^\alpha = \mu \epsilon^{\alpha\beta\gamma} \partial_\beta a_\gamma \tag{4.5}$$

This equation of motion demonstrates flux binding. To see this, let us look at the 0th component of this equation. We have

$$q j^0 = q \sum_{n=1}^{N} \delta(\mathbf{x} - \mathbf{x}_n) = \mu(\nabla \times \vec{a}) = \mu b$$

where we have defined a "Cherns-Simons" magnetic field to be the curl of the the Chern-Simons vector potential. In other words this equation attaches a delta function flux tube with flux q/μ at the position of each charge q. So we have achieved charge-flux binding!

We might expect that the phase obtained by exchanging two charges in this theory would be the charge times the flux or $\theta = q^2/\mu$. Actually, this is not right! The correct answer is that the statistical phase is

$$\theta = q^2/(2\mu)$$

. To see why this is the right answer, we can multiply our equation of motion 4.5 by a_α and then plug it back into the Lagrangian 4.3. We then end up with

$$\Delta \mathcal{L} = -\frac{q}{2} j^\alpha a_\alpha$$

In other words, the dynamics of the Chern-Simons vector potential itself cancels excatly half of the Lagrangian density, and hence will cancel half of the accumulated phase when we exchange two particles with each other!

If we are interested in calculating a propagator for our particles we can write

$$\sum_{\text{paths } \{\mathbf{x}\}} \sum_{\text{all } \mathrm{a}_\alpha(\mathrm{x})} e^{i(S_0 + S_{CS} + S_{coupling})/\hbar} \tag{4.6}$$

Here the first sum is the usual sum over paths that we have discussed above. The second sum is the sum over all possible configurations of the field a($\mathbf{x}$). Note this is all configurations in space and time so it is effectively a path integral for a field. (This is potentially everything you ever need to know about field theory!). Often the sum over field configurations is written as a functional integral

$$\sum_{\text{all } \mathrm{a}_\alpha(\mathrm{x})} \rightarrow \int \mathcal{D} a_\mu(x)$$

Formally when we write a functional integral we mean[12] that we should divide space and time into little boxes and within each box integrate over all possible values of a. Fortunately, we will not need to do this procedure explicitly.

We thus rewrite Eq. 4.6 as

$$\sum_{\text{paths } \{\mathbf{x}\}} e^{iS_0/\hbar} \int \mathcal{D} a_\mu(x)\ e^{iS_{CS}/\hbar}\ e^{i(q/\hbar)\int_{paths} dl^\alpha a_\alpha}$$

[11] It may look like the right result would be to have $\mu/2$ on the right hand side, given that it is $\mu/2$ in equation 4.3. However, note that when we differentiate with respect to a_α on the left hand side, we also generate an identical factor of $\mu/2$ and these two add up.

[12] Making strict mathematical sense of this type of integral is not always so easy!

where S_0 is the action of the particles following the path but not interacting with the gauge field, S_{CS} is the action of the Chern-Simons gauge field alone. The final term represents the coupling of the gauge field to the path of the particles — it is an integral that follows the path of the particles and integrates the vector potential along the path. This is precisely the phase accumulated by a particle in the vector potential. It is an example of a Wilson-line operator, which we will see again shortly.

Once the integration over the Chern-Simons field is done, we obtain

$$\sum_{\text{paths }\{\mathbf{x}\}} e^{iS_0/\hbar + i\theta W(path)}$$

where W is the winding number of the path and θ is the anyon statistical angle.

Something we have pointed out above is that the vacuum of an anyon theories knows about the statistics of the particles, even when the particles are not present (i..e, the ground state degeneracy on a torus matches the number of particle species). Thus, in the absence of particles, we will be interested in

$$Z(\mathcal{M}) = \int_{\mathcal{M}} \mathcal{D}a_\mu(x)\ e^{iS_{CS}/\hbar}$$

where $\mathcal{M}$ is the space-time manifold we are considering. In fact this integral will tell us about the ground state degeneracy (it is exactly the ground state degeneracy if we are considering a spatial manifold Σ, such as a torus, crossed with our time direction made periodic instead of infinite.) As we might expect, this quantity will be a topological invariant of the space-time manifold. That is, smooth deformations of $\mathcal{M}$ do not change its value. (See chapter appendix).

4.5 NonAbelian Chern-Simons theory: The paradigm of TQFT

Refs: As in section *** below, see articles by Atiyah, and Wittens seminal 1989 paper. See also the discussion in Nayak et al. Among 2+1D TQFTs pretty much everything of interest is somehow related to Chern-Simon theory – however, we don't generally have the luxury of working with Abelian theory.

We can generalize the Abelian Chern-Simons theory we discussed above by promoting the gauge field a_α to be not just a vector of numbers, but rather a vector of matrices.[13] More precisely we are making the vector potential take values in a Lie algebra[14] For example, if we choose to work with the Lie algebra of $SU(2)$ we can write

$$a_\mu(x) = a_\mu^k(x)\sigma_k \frac{i}{2}$$

where σ_k are the Pauli matrices. (the factor of 2 and i are a useful convention, although other conventions exist). Note that now that a_μ is matrix valued it becomes noncommutative and we have to be very careful about the order in which we write factors of a_μ.

The fundamental quantity that we should think about is the Wilson[15] loop operators

$$W_L = \mathrm{Tr}\left[P \exp\left(i\frac{q}{\hbar}\oint_L dl^\mu a_\mu\right)\right]$$

[13] If you have studied Yang-Mills theory, you already know about non-abelian vector potentials.

[14] A Lie Group is a group which is also a manifold. The group of complex phases of unit modulus is an example of this — it is also a circle. Another example, is the group $SU(2)$ (2 by 2 unitary matrices with determinant 1) which is equivalent to the three sphere. The definition of a three sphere is the set of all real 4-vectors (x, y, z, w) of unit length $x^2 + y^2 + z^2 + w^2 = 1$. Every element of $SU(2)$ can be written as $ix\sigma_x + iy\sigma_y + iz\sigma_z + w\mathbf{1}$ where σ_i are the Pauli matrices and $\mathbf{1}$ is the identity matrix (check that this does indeed give a unitary unit determinant matrix!). A Lie *algebra* is the algebra of infinitesimal changes in this group. We can think of this as writing all the elements of the group close to the identity as $\mathbf{1} + \epsilon\mathbf{g}$ where ϵ is very small, then the elements g are the Lie algebra of $SU(2)$. Thus the algebra is spanned elements $i\sigma_k$ in this case. The groups $SU(N)$ and $SO(N)$ are also Lie groups.

[15] These are named for Ken Wilson, who won a Nobel Prize for his work on the renormalization group and critical phenomena.

where here the integral follows some path L. (The trace can be taken in any representation, although it is perhaps most useful to stick with the fundamental representation of the group.). Here, the P symbol indicates path ordering — analogous to the usual time ordering of quantum mechanics. The issue here is that $a_\mu(x)$ is a matrix, so when we try to do the integral and exponentiate, the $a_\mu(x)$ and $a_\mu(x')$ do not commute. The proper interpretation of the path orderered integral is then to divide the path into tiny pieces of length dl. We then have (setting $q = \hbar = 1$ for simplicity of notation)

$$P\exp(i\oint_L dl^\mu a_\mu) = [1 + ia_\mu(x_1)dl^\mu(x_1)]\ [1 + ia_\mu(x_2)dl^\mu(x_2)]\ [1 + ia_\mu(x_3)dl^\mu(x_3)]\ \ldots \tag{4.7}$$

where $x_1, x_2, x_3, \ldots$ are the small steps along the path. Since the exponent is an integral of a vector potential around a loop, we can think of this path ordered integral as giving us $e^{iq\Phi/\hbar}$ where Φ is now some sort of matrix valued flux.

The proper gauge transformation in the case of a nonabelian gauge field is given by

$$a_\mu \to U a_\mu U^{-1} - i\frac{\hbar}{q} U\partial_\mu U^{-1} \tag{4.8}$$

Where $U(x)$ is a matrix (which is a function of position and time) which acts on the matrix part of a_μ. Note that this is just the nonabelian analogue of the gauge transformation in Eq. 4.4 (take $U = e^{i\chi}$). To see that this gauge transformation leaves invariant the Wilson loop operators, see Appendix.

The Chern-Simons action is now written as

$$S_{CS} = \frac{\hbar k}{4\pi}\int_{\mathcal{M}} d^3x\ \epsilon^{\alpha\beta\gamma}\ \mathrm{Tr}\left[a_\alpha\partial_\beta a_\gamma - \frac{2i}{3}a_\alpha a_\beta a_\gamma\right]$$

Note that the second term in the brackets would be zero if the a_α were commutative (as in the Abelian case above, where we have no such term!).

The Chern-Simons action is not quite gauge invariant. If in Eq.4.8 we use a unitary matrix which is "close" to the identity (i.e., can be continuously deformed to the identity) then we will find the action is gauge invariant up to a surface term (which vanishes for a closed manifold). To see this we can simply use

$$U = \exp(i\epsilon H(x)) = 1 + i\epsilon H(x) \tag{4.9}$$

for a hermitian matrix $H(x)$ (which again can be a function of space and time) where we can assume ϵ is very small. Substituting this in to Eq. 4.8 and then into the Chern-Simons action, we find that (to lowest order in ϵ) the action is indeed gauge invariant (See appendix). However, it turns out that the unitary function of space and time $U(x)$ has topologically disconnected components — that is, we cannot continuously reach some functions $U(x)$ starting from the identity and making small deformations. This is not immediately obvious, but it turns out that under such "large" gauge transformations, we have

$$S_{CS} \to S_{CS} + 2\pi kn\hbar$$

for some integer n which is some sort of winding number[16] of the function $U(x)$. This may look problematic, but we note that the only thing entering our functional integral is $e^{iS_{CS}/\hbar}$, not the Chern-Simons action itself. Thus, so long as we choose k, the so-called "level", as an integer, then we have a well defined functional integral of the form

$$Z(\mathcal{M}) = \int_{\mathcal{M}} \mathcal{D}a_\mu(x)\ e^{iS_{CS}/\hbar}$$

where the result $Z(M)$ turns out to be a manifold invariant (see chapter appendix).

[16]In the case of the gauge group being $SU(2)$, as mentioned in footnote 14, the gauge group is isomorphic to the manifold S^3. So if the manifold happens to be S^3 then we are looking at mappings from S^3 back into S^3. The mapping of Eq. 4.9 corresponds to zero winding number (can be continuously deformed to $U = 1$ everywhere). However, we also can consider the identity mapping that S^3 maps into S^3 in the obvious way (each point goes to the same point) which gives an $n = 1$ mapping (a 1-to-1 mapping). One can also construct 2-to-1 mappings which have winding $n = 2$ etc.

Figure 4.9: A cartoon of a 3 manifold with a link embedded in it.

.

The insertion of the Wilson loop operator into the path integral gives a knot invariant of the link L that the Wilson loop follows[17]. Often we will think about our link as being embedded in a simple manifold like the three sphere[18], which we denote as S^3.

So for example, to find the link invariant corresponding to the two linked strings in Fig. 4.9, we have

$$\text{Knot Invariant} = \frac{Z(S^3, L_1, L_2)}{Z(S^3)} = \frac{\int_{S^3} \mathcal{D}a_\mu(x)\ W_{L1}W_{L2}\ \ e^{iS_{CS}/\hbar}}{\int_{S^3} \mathcal{D}a_\mu(x)\ e^{iS_{CS}/\hbar}}$$

And indeed, if we choose to work with the gauge group $SU(2)$ at level k we obtain the Kauffman invariant of the knot with $A = -(-i)^{(k+1)/(k+1)}$.

4.6 Appendix: Odds and Ends about Chern Simons Theory

4.6.1 Gauge Transforms with Nonabelian Gauge Fields

Let us define a Wilson-line operator, similar to the Wilson loop but not forming a closed loop, i.e., going along a curve C from space-time point x to point y.

$$W_C(x,y) = \text{Tr}\left[P \exp\left(i\frac{q}{\hbar}\int_C dl^\mu a_\mu\right)\right]$$

Under a gauge transformation function $U(x)$ we intend that the Wilson line operator transform as

$$W_C(x,y) \to U(x)\ W_C(x,y)\ U(y)^{-1}$$

Clearly this obeys composition of paths, and will correctly give a gauge invariant result for a closed Wilson loop.

Now to see what is required for the gauge field a_μ such that this works, we consider

$$W_C(x, x+dx) = 1 + i\frac{q}{\hbar}a_\mu dx^\mu$$

[17]The observant reader will note that we have not specified the "framing" of the knot – i.e, if we are to think of the world-line as being a ribbon not a line, we have not specified how the ribbon twists around itself. In field theory language this enters the calculation by how a point-splitting regularization is implemented.

[18]The three sphere we can imagine as being $\mathbb{R}^3$ with a point added at infinity, the same way we can think of a 2 sphere as being a plane with a point adeed at infinity. Another way to think about the 3-sphere is as the surface of a 4-ball B^4. If we can imagine 4-dimensional space with real coordinates (x, y, z, w), the 3-sphere is the set of points such that $x^2 + y^2 + z^2 + w^2 = 1$. Oops, just realized I gave this definition in the previous footnote too!

and its transformation should be

$$\begin{aligned} W_C(x, x+dx) \to U(x) W_C(x, x+dx) U(x+dx)^{-1} &= U(x)[1 + i\frac{q}{\hbar} a_\mu dx^\mu] U(x+dx)^{-1} \\ &= U(x)[1 + i\frac{q}{\hbar} a_\mu dx^\mu][U(x)^{-1} + dx^\mu \partial_\mu U(x)^{-1}] \\ &= 1 + i\frac{q}{\hbar}[U a_\mu U^{-1} - i\frac{\hbar}{q} U \partial_\mu U^{-1}] \end{aligned}$$

Thus the gauge transform rule Eq. 4.8 correctly gives a gauge invariant Wilson loop operator.

4.6.2 Gauge Invariance of Chern-Simons action under small transforms

Just plug it in!

4.6.3 Chern Simons Action is Metric Independent

You will often see books state that you don't see the metric $g_{\mu\nu}$ written anywhere — QED. But that kind of misses the point!

A differential geometer would see that one can write the Chern-Simons action in differential form notation

$$S_{CS} = \frac{k}{4\pi} \int (a \wedge da + \frac{2}{3} a \wedge a \wedge a)$$

which then makes it "obvious" that this is metric independent.

In more detail however, we must first declare how the gauge field transforms under changes of metric. It is a "1-form" meaning it is meant to be integrated along a line to give a reparameterization invariant result, such as in the wilson loops

$$\int da = \int dx^\mu a_\mu(x) = \int dx'^\mu \frac{\partial x^\mu}{\partial x'^\nu} a_\nu(x')$$

This means that under reparameterization $x'(x)$ we have

$$a_\mu(x) = \frac{\partial x^\mu}{\partial x'^\nu} a_\nu(x')$$

such that the line integral remains invariant under a reparameterization of the space.

Now, if we make this change on all of the a's in the the Chern-Simons action we obtain

$$\begin{aligned} &\epsilon^{\alpha\beta\gamma} \ \mathrm{Tr}\left[a_\alpha \partial_\beta a_\gamma - \frac{2i}{3} a_\alpha a_\beta a_\gamma\right] \to \\ &\epsilon^{\alpha\beta\gamma} \frac{\partial x^\alpha}{\partial x'^{\alpha'}} \frac{\partial x^\beta}{\partial x'^{\beta'}} \frac{\partial x^\gamma}{\partial x'^{\gamma'}} \mathrm{Tr}\left[a_{\alpha'} \partial_{\beta'} a_{\gamma'} - \frac{2i}{3} a_{\alpha'} a_{\beta'} a_{\gamma'}\right] \end{aligned}$$

But notice that the prefactor, including the ϵ this is precisely the Jacobian determinant and can be rewritten as

$$\epsilon^{\alpha'\beta'\gamma'} \det[\partial x / \partial x']$$

Thus the integral can be changed to the dx' variables and the form of the integral is completely unchanged.

In fact, this feature of the Chern-Simons Lagrangian is fairly unique. Given that we have a single gauge field $a_\mu(x)$ this is the *ONLY* (3-form) gauge invariant Lagrangian density we can write down which will give a topological invariant!

4.6.4 Framing of the Manifold — or doubling the theory

There is a bit of a glitch in Chern-Simons theory. We want the Chern-Simons functional $Z(\mathcal{M})$ to be a function of the topology of $\mathcal{M}$ only. This is *almost* true — it is true up to a phase. In order to get the phase, you need to specify one more piece of information which can be provided in several ways (often called a 2-framing). This additional piece of information is most easily described by saying that you need to specify a bit of information about the topology of the 4-manifold $\mathcal{N}$ that $\mathcal{M}$ bounds $\mathcal{M} = \partial\mathcal{N}$. It is a fact that all closed 3-manifolds are the boundary of some 4-manifold — in fact, of many possible 4-manifolds. The phase of $Z(\mathcal{M})$) is senstive only to the so-called "signature" of the 4-manifold $\mathcal{N}$. (Consult a book on 4 manifold topology if you are interested!)

The fact that the Chern-Simons theory should depend on some information about the 4-manifold that $\mathcal{M}$ bounds may sound a bit strange. It is in fact a sign that the Chern-Simons theory is "anomolous". That is, it is not really well defined in 3-dimensions. If you try to make sense of the functional integral $\int \mathcal{D}a_\mu$, you discover that there is no well defined limit by which you can break up space-time into little boxes and integrate over a in each of these boxes. However, if you extend the theory into 4-dimensions, then the theory becomes well behaved. This is not unusual. We are familiar with lots of cases of this sort. Perhaps the most famous example is the fermion doubling problem. You cannot write down a theory for a single chirality fermion in d dimensions without somehow getting the other chirality. However, you can think of a system extended into $d+1$ dimensions where one chirality ends up on one of the d-dimensioanl boundaries and the other chirality ends up on the other d dimensional boundary.

So to make Chern-Simons theory well defined, you must either extend into 4d, or you can "cancel" the anomoly in 3d by, for example, considering two, opposite chirality Chern-Simons theories coupled together (so-called "doubled" Chern-Simons theory). The corresponding manifold invariant of a doubled theory gets $Z(\mathcal{M})$ from the righthanded theory and its complex conjugate from the left handed thoery, thus giving an end result of $|Z(\mathcal{M})|^2$ which obvioulsy won't care about the phase anyway!

4.6.5 Chern Simons Canonical Quantization for the Abelian Case

One can consider the Chern-Simons theory as a quantum mechanical theory with wavefunctions and operators (i.e., not in path integral language). To do this, we need to find the commutation relations. Note in the Chern-Simons Lagrangian terms like $\partial_0 a_y$ multiply a_x and vice versa. This means that $a_y(x)$ is the momentum conjugate to $a_x(x)$ and vice versa. We thus have the commutation relations

$$[a_x(\vec{x}), a_y(\vec{x}')] = \frac{i\hbar}{\mu}\delta(\vec{x} - \vec{x}')$$

The arguments $\vec{x}$ here live in 2 dimensions. Consider now the wilson loop operators around the two differnt handles

$$W_j = \exp(i(q/\hbar) \oint_L \vec{dl} \cdot \vec{a})$$

where here j indicates we have a loop around either cycle 1 or cycle 2 of our torus. The two paths must intersect at one point and therefore, due to the above commutations, do not commute with each other. We can use the indentity that

$$e^A e^B = e^B e^A e^{[A,B]}$$

which holds when $[A, B]$ is a number not an operator. This the gives us

$$W_1 W_2 = e^{iq^2/\mu\hbar} W_2 W_1 = e^{i\theta} W_2 W_1$$

where θ is the statistical angle of the theory. Thus the Wilson loop operators act just like operators T_1 and T_2 which created particle-hole pairs and moved them around the handle then reannihilated. So even without discussing particles, the ground state wavefunction of the Chern-Simons theory is degenerate!

Chapter 5

Short Digression on Quantum Gravity

5.0.1 Why this is hard

Little is known about quantum gravity with any certainty at all. What we do know for sure is the value of some of the fundamental constants that must come into play. We expect the gravitational constant G, the speed of light c and of course Planck's constant $\hbar$. From these we can put together an energy scale, known as the Planck Scale

$$E = \sqrt{\frac{\hbar c^5}{G}} \approx 10^{28}\,\text{eV}$$

The temperature of the world around us is about .03eV. Chemistry, visible light, and biology occur on the scale of 1eV. The LHC accelerator probes physics on the scale of roughly $10^{13}eV$. This means trying to guess anything about the Planck scale is trying to guess physics on an energy scale 15 orders of magnitude beyond what any accelerator[1] experiment has ever probed! We must surely accept the difficult possibility that any physical principle we hold dear from all of our experiments on low energy scales could no longer hold true at the Planck scale! The only thing that is really required is that the effective low energy theory matches that which we can see at the low energies in the world around us!

5.0.2 Which Approach?

There are several approaches to quantum gravity. While I will not make any statement about which approaches is promising, and which approaches are crazy and overpublicized[2], I am comfortable stating that many of these investigations have led to increadibly interesting and important things being discovered. While in some cases (maybe in most cases) the discoveries may be more about math than about physics, they are nonetheless worthwhile investigations that I support.

5.1 Some general principles?

We have to choose general principles that we want to believe will always hold, despite the fact that we are considering scales of energy and length 14 orders of magnitude away from anything we have

[1] Cosmic ray observations have been made at several orders of magnitude higher still — but very little can be deduced from these extremely rare and uncontrolled events. A famous event known as the "oh my god particle" was apparently $10^{20}eV$, still 8 orders of magnitude away from the Planck scale.

[2] For information on the wars between some of the different approaches to quantum gravity, see "The String Wars", or "The Trouble With Physics" or "Not Even Wrong".. or see repsonses to these such as the article by Polchinski in the American Scientist, or if you are ready for a major rant, the response by Lubos Motl.

ever observed or measured. Much of the community feels that the most fundamental thing to hold onto is the Feynman picture of quantum mechanics — that all space-time histories must be allowed. We might write a quantum partition function of the form

$$Z = \sum_{\text{All universes}} e^{iS/\hbar} \tag{5.1}$$

where the sum is now over everything that could happen — it is the ultimate sum over all histories! Obviously such a thing is hard to even contemplate. Several key simplifications will make contemplation easier

(1) Lets ignore matter. Let us (at least to begin with) try to model only universes which are completely devoid of substance and only contain vacuum. Thus the universe contains only the space-time metric. Doing this, the Einstein-Hilbert action[3] for gravity takes the form

$$S_{Einstein} \sim \int_{\mathcal{M}} dx\ R\sqrt{-g}$$

where the integration is over the entire space-time manifold $\mathcal{M}$, and g is the space-time metric and R is the Ricci scalar. One might imagine that we could construct a real theory of quantum gravity by plugging the Einstein-Hilbert action into the path integral form of Eq.5.1. We obtain

$$Z = \int \mathcal{D}g\ e^{iS_{Einstein}(g)/\hbar}$$

Even without matter in the universe, the model is very nontrivial because the space-time metric can fluctuate — which we know very well from the 2015 measurment of gravitational waves! Even in this limit no one has fully made sense of this type of path integral without many additional assumptions.

(2) Let us simplify even more by considering a 2+1 dimensional universe.

For a huge amount of information on 2+1 dimensional quantum gravity, see http://www.livingreviews.org/lrr-2005-1. For the key parts that are interesting to us, see the article by Witten on "2+1 Dimensional Gravity as an Exactly Solvable System."

We are used to the idea that many things simplify when we go to lower dimension. Indeed, that is what happens here. In 2+1 dimension, there is an enormous simplification that there are no gravity waves! Why not? In short, there are just not enough degrees of freedom in a 2+1 dimensional metric to allow for gravity waves. (For more information on this fact see the appendix to the chapter.) As a result, the only classical solution of the Einstein equations in the vacuum is that $R = 0$ and that is all! I.e., the universe is flat and there are no fluctuations. (One can also have a cosmological constant in which case $R = 2\Lambda g$ is the solution).

One might think that this means that gravity in 2+1 D is completely trivial. However, it is not. The space-time manifold, although everywhere curvature free, still has the possibility of having a *nontrivial topology*. Thus what we are interested in is actually the different topologies that our space-time manifold might have!

We thus rewrite Eq. 5.1 as

$$\begin{aligned} Z &= \sum_{\text{manifolds}\,\mathcal{M}} \int_{\mathcal{M}} \mathcal{D}g\ e^{iS(g)/\hbar} \\ &= \sum_{\text{manifolds}\,\mathcal{M}} Z(\mathcal{M}) \end{aligned}$$

where $S(g)$ is the Einstein-Hilbert action for a flat universe with metric g, the sum is over all different topologies of manifolds the universe might have, and the integration $\mathcal{D}g$ is an integration over all metrics subject to the condition that the manifold's topology is fixed to be $\mathcal{M}$.

[3]Written down by Hilbert first in 1915

Why would we be interested in such a quantity? In short, suppose we know what the topology is of our (d-dimensional universe) at a fixed time t. We want to know the amplitudes that the topology changes as t develops. I.e., is the space-time manifold of our universe of the form $\mathcal{M} = \Sigma \times$ time or does the space-time manifold split? For example, see the diagram 5.1.

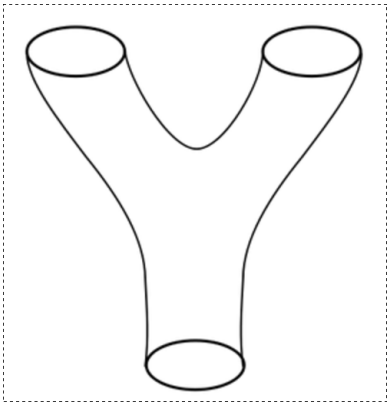


Figure 5.1: A manifold where the topology of a time-like slice changes as time progresses.

OK, here is the kicker: The rather surprising result is that the function $Z(\mathcal{M})$ is precisely the Chern-Simons partition function discussed above for an appropriately chosen gauge group!

5.2 Appendix: No Gravity Waves in 2+1 D

The short argument for this is as follows (stolen from the article by Carlip)

> In n dimensions, the phase space of general relativity is parametrized by a spatial metric at constant time, which has $n(n-1)/2$ components, and its conjugate momentum, which adds another $n(n-1)/2$ components. But n of the Einstein field equations are constraints rather than dynamical equations, and n more degrees of freedom can be eliminated by coordinate choices. We are thus left with $n(n-1) - 2n = n(n-3)$ physical degrees of freedom per spacetime point. In four dimensions, this gives the usual four phase space degrees of freedom, two gravitational wave polarizations and their conjugate momenta. If $n = 3$, there are no local degrees of freedom.

To put a bit more detail on this argument, if we write the flat metric as $\eta_{\mu,\nu} = \text{diag}[-1, 1, 1, \ldots]$ in any dimension, and we consider small deviations from a flat universe we have $g = \eta + h$, we can construct the trace-reversed

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}\eta^{\rho\sigma}h_{\rho\sigma}$$

In any dimension, gravitational waves in vacuum take the form

$$\bar{h}^{\mu\nu}{}_{,\nu} = 0$$

and

$$\Box\bar{h}_{\mu\nu} = 0$$

where the comma notation indicates derivatives, and indices are raised and lowered with η.

In any dimension we will have the gravitational wave of the form

$$\bar{h}_{\mu\nu} = \epsilon_{\mu\nu}e^{ik^\rho x_\rho}$$

where the polarization is $\epsilon_{\mu\nu}$ is orthogonal to the lightlike propagation wavevector, $k^\mu k_\mu = 0$.

$$\epsilon_{\mu\nu} k^\nu = 0 \tag{5.2}$$

However, one must also worry about gauge freedoms. We can redefine our coordinates and change the form of the metric without changing any of the spatial curvatures. In particular making a coordinate transform $x \to x' - \xi$, we have

$$\bar{h}_{\mu\nu} \to \bar{h}_{\mu\nu} - \xi_{\nu,\mu} - \xi_{\mu,\nu} + \eta_{\mu,\nu}\xi^{\alpha}_{,\alpha}$$

Now here is the key. In 2+1 D for *any* matrix ϵ you choose, you can always find a

$$\xi_\mu = A_\mu e^{ik^\rho x_\rho}$$

such that

$$\bar{h}_{\mu\nu} = \epsilon_{\mu\nu} e^{ik^\rho x_\rho} = \xi_{\nu,\mu} - \xi_{\mu,\nu} + \eta_{\mu,\nu}\xi^{\alpha}_{,\alpha}$$

This means that the wave is pure gauge, and the system remains perfectly flat! I.e., if you calculate the curvature with this form of $\bar{h}$, you will find zero curvature.

To be more precise, we find

$$\epsilon_{\mu,\nu} = A_\mu k_\nu - A_\nu k_\mu + \eta_{\mu\nu} A^\sigma k_\sigma$$

And any ϵ that satisfies Eq. 5.2 can be represented with some vector A. It is easy to check this by counting degrees of freedom. ϵ has 6 degrees of freedom in 2+1D, but Eq. 5.2 is 3 constraints, and A has three parameters, so we should always be able to solve the equation for A given ϵ.

Chapter 6

Topological Quantum Field Theory

We already have a rough picture of a Topological Quantum Field Theory (TQFT) as a quantum theory that depends on topological properties not on geometric properties. (For example, it matters that particle 1 traveled around particle 2, but it doesn't matter how far apart they are.)

We can formalize these ideas by saying that the theory should be independent of small deformations of the space-time metric. We might say that

$$\frac{\delta}{\delta g_{\mu\nu}}\langle \text{any correlator}\rangle = 0$$

This is a completely valid way to define a TQFT. However, a somewhat different definition has been given by a set of Axioms by Atiya which are in some sense much more informative.

6.1 Paraphrasing of Atiyah's Axioms

Several good references for this. Article by Atiyah. Book by Atiyah. Article by witten.

Here I'm going to give a rough interpretation of Aityah's axioms for a TQFT. To begin with we will consider space-time manifolds with no particle in them. Later on we will discuss adding particles and moving them around in space-time too.

We will consider a $d+1$ dimensional space-time manifold which we call $\mathcal{M}$, and d dimensional oriented slice Σ – we should think of this slice as being the space at a fixed time. Almost always we will be thinking of $d = 2$, although the axioms are quite general and can be applied to any d.

AXIOM 1: A d-dimensional space Σ is associated with a Hilbert space $V(\Sigma)$ which depends only on the topology of Σ.

We call the space V-for vector space, although sometimes people call it H for Hilbert space.

For example, we have seen that if Σ is a torus, there is a nontrivial Hilbert space coming from the ground state degeneracy.

Note that when we add particles to the system (and we have not done this yet), if the particles are non-abelian, then there will be a Hilbert space associated with the degeneracy that comes with such non-abelian particles.

AXIOM 2: the disjoint union of two d-dimensional spaces Σ_1 and Σ_2 will be associated with a Hilbert space which is the tensor product of the Hilbert spaces associated with each space. I.e.,

$$V(\Sigma_1 \cup \Sigma_2) = V(\Sigma_1) \otimes V(\Sigma_2)$$

In particular this means that the vector space associated with the null or empty space $\emptyset$ must be just the complex numbers, or

Implies:

$$V(\emptyset) = \mathbb{C}$$

Because $\emptyset \cup \Sigma = \Sigma$ and $\mathbb{C} \otimes V(\Sigma) = V(\Sigma)$ so the result follows[1].

AXIOM 3: If $\mathcal{M}$ is a $(d+1)$-dimensional manifold with boundary $\Sigma = \partial M$, then we associate a particular element of the vector space $V(\Sigma)$ with this manifold. We write

$$Z(\mathcal{M}) \in V(\partial\mathcal{M})$$

where the association, i.e, which state in the vector space is chosen, again depends only on the topology of $\mathcal{M}$.

Here we should think of ∂M as being the space-like slice of the system at a fixed time, and $V(\partial\mathcal{M})$ as being the possible Hilbert space of ground states. The rest of $\mathcal{M}$ is the space-time history of the system, and $Z(\mathcal{M})$ is the particular wavefunction that is picked out.

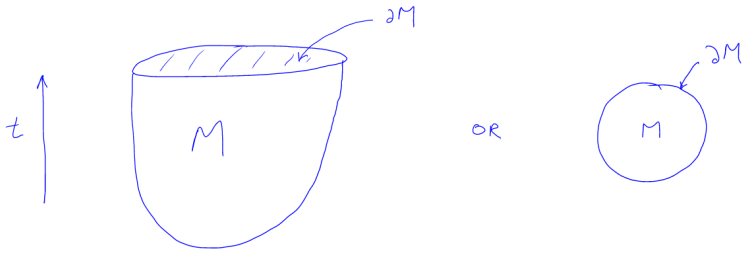


Figure 6.1: Two depictions of a space-time manifold $\mathcal{M}$ with boundary $\partial\mathcal{M}$. The left depiction is problematic because the only boundary of the manifold is supposed to be the top surface ∂M. The right depiction is more accurate in this sense, although it depicts a 2D $\mathcal{M}$ and 1D $\partial\mathcal{M}$

The point of this axiom is to state that the particular wavefunction of a system $Z(\mathcal{M})$ which is chosen from the available vector space depends on the space-time history of the system. We have seen this principle before several times. For example, we know that if a particle-antiparticle pair is taken around a handle, this changes which wavefunction we are looking at – this process would be part of the space-time history.

Implies: For $\mathcal{M}$ closed, we have $\partial M = \emptyset$, the empty space, so

$$Z(\mathcal{M}) \in \mathbb{C}$$

i.e., the TQFT must assign a manifold a topological invariant which is a complex number.

AXIOM 4: Reversing Orientation

$$V(\Sigma^*) = V^*(\Sigma)$$

where by Σ^* we mean the same surface with reversed orientation, whereas by V^* we mean the dual space — i.e., we turn bras into kets. It is a useful convention to keep in mind that the orientiation of the normal of $\partial\mathcal{M}$ should be pointing out of $\mathcal{M}$. See Fig. 6.2

GLUING: If we have two manifolds $\mathcal{M}$ and $\mathcal{M}'$ which have a common boundary $\partial\mathcal{M} = (\partial\mathcal{M})^*$ we can glue these two manifolds together by taking inner products of the corresponding states as shown in Fig. 6.3. Here we have $\Sigma = \partial\mathcal{M} = (\partial\mathcal{M}')^*$ so we can glue together the two manifolds along their common boundary to give

$$Z(\mathcal{M} \cup_\Sigma \mathcal{M}') = \langle Z(\mathcal{M}')|Z(\mathcal{M})\rangle$$

[1]If this sounds confusing, remember the space $\mathbb{C}$ is just the space of length 1 complex vectors, and tensoring a length n vector with a length m vector gives a size n by m matrix, so tensoring a vector of length n with a length 1 vector gives back a vector of length n.

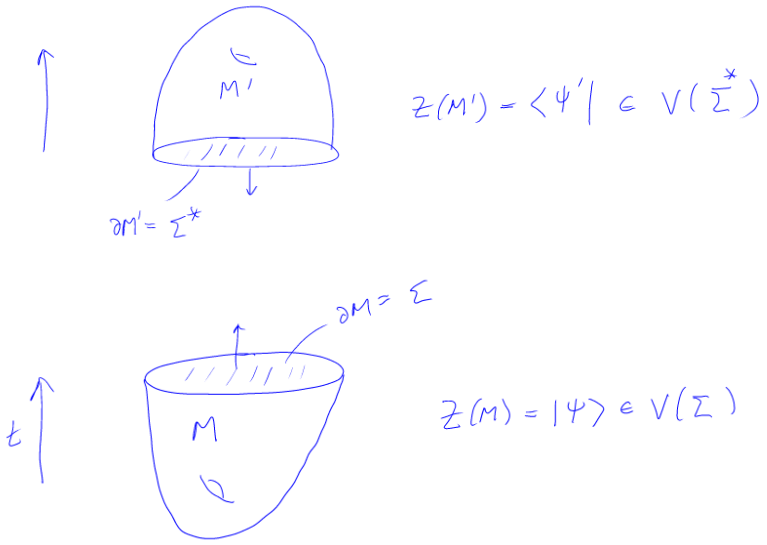


Figure 6.2: In this picture $\mathcal{M}$ and $\mathcal{M}'$ are meant to fit together since they have a common boundary but with opposite orientation $\Sigma = \partial\mathcal{M} = \partial\mathcal{M}'^*$. Here $\langle\psi'| = Z(\mathcal{M}') \in V(\Sigma^*)$ lives in the dual space of $|\psi\rangle = Z(\mathcal{M}) \in V(\Sigma)$

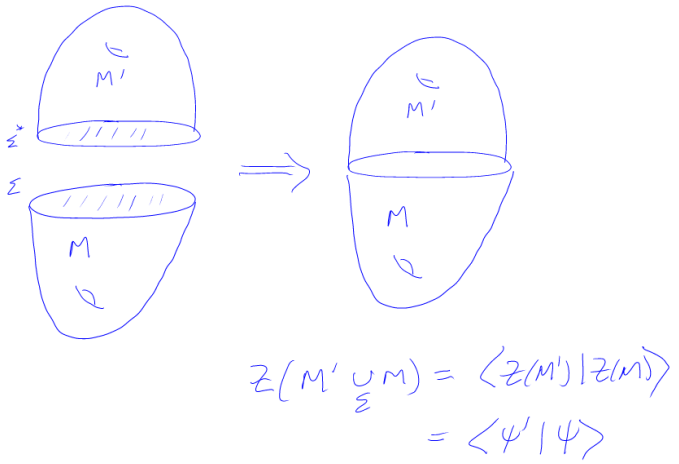


Figure 6.3: Gluing two manifolds together by taking the inner product of the wavefunctions on their common, but oppositely oriented, boundaries.

COBORDISM: Two manifolds Σ_1 and Σ_2 are called "cobordant" if their disjoint union is the boundary of a manifold $\mathcal{M}$.

$$\partial\mathcal{M} = \Sigma_1 \cup \Sigma_2$$

(or we say that $\mathcal{M}$ is a cobordism between Σ_1 and Σ_2. See Fig. 6.4. We thus have $Z(\mathcal{M}) \in V(\Sigma_1^*) \otimes V(\Sigma_2)$, so that we can write

$$Z(\mathcal{M}) = \sum_{\alpha\beta} U^{\alpha\beta} \, |\psi_{\Sigma_2,\alpha}\rangle \otimes \langle\psi_{\Sigma_1,\beta}|$$

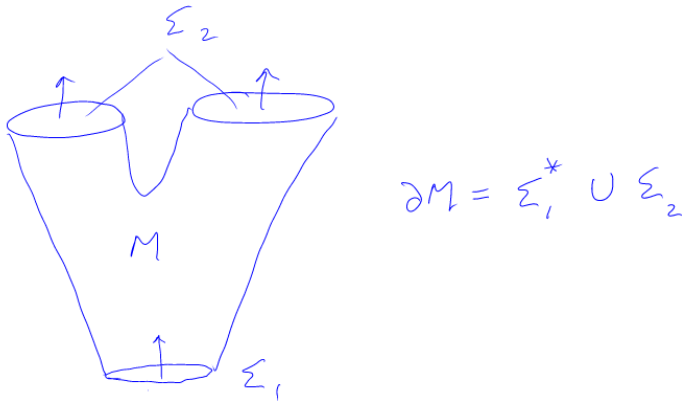


Figure 6.4: $\mathcal{M}$ is the cobordism between Σ_1^* and Σ_2. I.e., $\partial\mathcal{M} = \Sigma_1^* \cup \Sigma_2$. Note that we have reversed orientation of Σ_1 here.

where $|\psi_{\Sigma_2,\alpha}\rangle$ is the basis of states for $V(\Sigma_2)$ and $\langle\psi_{\Sigma_1,\beta}|$ is the basis of states for $V(\Sigma_1^*)$. We can thus think of the cobordism $\mathcal{M}$ as being the unitary evolution between the vector spaces $V(\Sigma_1)$ and $V(\Sigma_2)$ (here U is unitary).

IDENTITY COBORDISM: If we have $\mathcal{M} = \Sigma \times I$ where I is the one dimensional interval (We could call it the 1-disk, D^1 also, perhaps a better notation?) then the boundaries as Σ and Σ^* (See fig.6.5, and the cobordism implements a map between $V(\Sigma)$ and $V(\Sigma)$. Since the interval can be topologically contracted to nothing, we can take this map to be the identity, see fig. 6.5.

$$Z(\Sigma \times I) = \sum_\alpha |\psi_{\Sigma,\alpha}\rangle \otimes \langle\psi_{\Sigma,\alpha}| = \text{identity}$$

We can now consider taking the top of the interval I and gluing it to the bottom to construct a closed manifold $\mathcal{M} = \Sigma \times S^1$, where S_1 means the circle (or 1-sphere). See figure 6.5. We then obtain

$$Z(\Sigma \times S^1) = \text{Tr}\left[\sum_\alpha |\psi_{\Sigma,\alpha}\rangle \otimes \langle\psi_{\Sigma,\alpha}|\right] = \text{Dim}[V(\Sigma)] \tag{6.1}$$

We simply obtain the dimension of the Hilbert space $V(\Sigma)$, or in other words, the ground state degeneracy of the 2-manifold Σ.

As we have discussed above for the torus T^2 we have

$$\text{Dim}\, V(T^2) = \#\text{ of particle species} \tag{6.2}$$

which we argued based on non-commutativity of taking anyons around the handles of the torus. On the other hand, for a 2-sphere S^2, we have

$$\text{Dim}\, V(S^2) = 1 \tag{6.3}$$

since there are no noncontractable loops. (We will justify these statments more later!)

6.2 Adding Particles

We now consider extending the ideas of TQFT to space-time manifolds with particle world-lines in them. This is sometimes called a "relative" TQFT, as compared to the case with no particles, which is sometimes called an "absolute" TQFT.

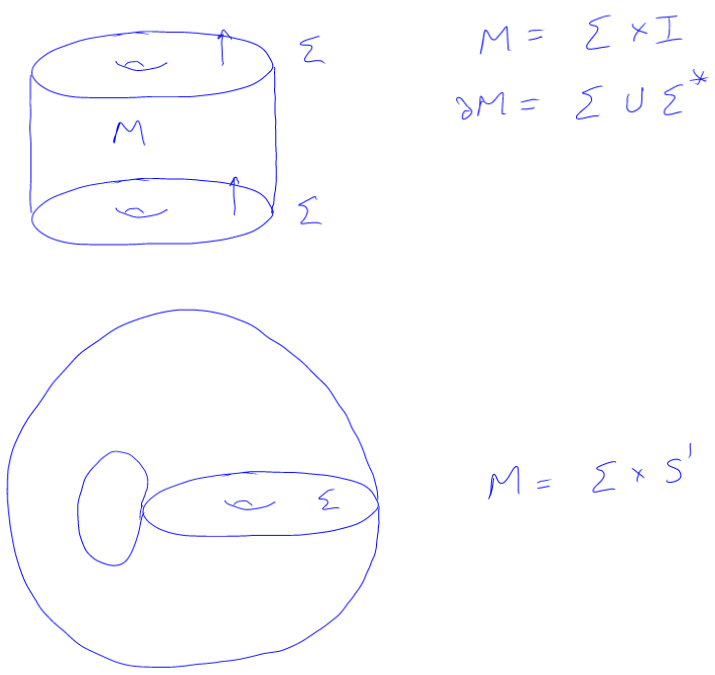


Figure 6.5: Top: A cobordism that can be topologically contracted to nonthing acts as the identity on the Hilbert space $V(\Sigma)$. Bottom: Gluing the top of $\Sigma \times I$ to the bottom we obtain $\mathcal{M} = \Sigma \times S^1$. The important fact is that $Z(\Sigma \times S^1$ is just the ground state degeneracy of Σ as a 2-manifold.

Figure 6.6: A 2-manifold with particles in it, which are marked and labeled points. We now call the combination (the manifold and the marked points) Σ for brevity.

At any rate, let us imagine there are different particle types which we can label as a, b, c, and so forth. We now imagine a 2-manifold with some marked and labeled points as shown in Fig. 6.6. We call the combination of the 2-manifold with the marked points Σ for brevity. As with the case we studied above, Σ is associated with a Hilbert space $V(\Sigma)$. The dimension of this Hilbert space depends on the number and type of particles in the manifold (as we expect for non-abelian particles, the dimension will grow exponentially with the number of particles). We can span the space $V(\Sigma)$ with some basis states $|\psi_\alpha\rangle$ which will get rotated into each other if we move the marked points around within the manifold (i.e., if we braid the particles around each other).

Similarly a 3-manifold $\mathcal{M}$ is now supplemented with labeled links indicating the world lines of the particles. The world-lines should be directed unless the particles are their own antiparticles. The world lines are allowed to end on the boundary of the manifold $\partial\mathcal{M}$. See Fig. 6.7. Analogously we may sometimes call the combination of the manifold with its world lines $\mathcal{M}$, although sometimes we will write this as $\mathcal{M}; L$ where L indicates the "link" (or knot) of the world lines. As in the above

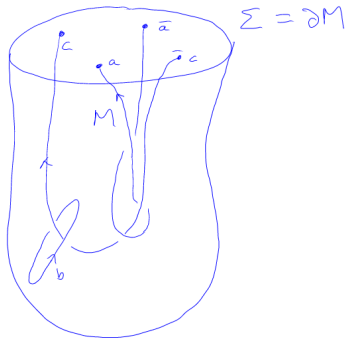


Figure 6.7: A 3-manifold $\mathcal{M}$ with particles in it, which are marked and labeled lines (the lines should be directed unless the particle is its own antiparticle). These world lines may end on the boundary $\Sigma = \partial\mathcal{M}$. The wavefunction on the boundary $\partial\mathcal{M}$ is determined by the spacetime history given by $\mathcal{M}$.

discussion, the spacetime history specifies exactly which wavefunction

$$|\psi\rangle = Z(\mathcal{M}) \in V(\partial\mathcal{M})$$

is realized on the boundary $\Sigma = \partial\mathcal{M}$.

We can now think about how we would braid particles around each other. To do this we glue another manifold $\mathcal{M}'$ to $\partial\mathcal{M}$ to continue the time evolution, as shown in Fig. 6.8. The final wave-

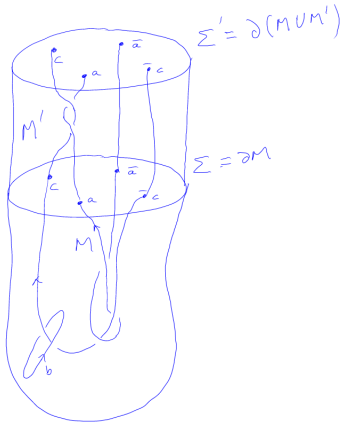


Figure 6.8: $\mathcal{M}'$ evolves the positions of the particles in time. Note that by $\mathcal{M}'$ we mean not just the manifold, but the manifold along with the world-lines in it.

function is written as

$$|\psi'\rangle = Z(\mathcal{M} \cup \mathcal{M}') \in V(\Sigma')$$

If we put the positions of the particles in Σ' at the same positions as the particles in Σ, then the Hilbert spaces, $V(\Sigma')$ is the same as $V(\Sigma)$, and we can think of $Z(\mathcal{M}')$ as giving us a unitary transformation on this Hilbert space – which is exactly what we think of as nonabelian statistics. We can write explicitly the unitary transformation

$$Z(\mathcal{M}') = \sum_{\alpha\beta} U^{\alpha\beta} \, |\psi_{\Sigma',\alpha}\rangle \otimes \langle\psi_{\Sigma,\beta}|$$

Note that if the particles stay fixed in their positions (or move in topologically trivial ways) then $\mathcal{M}'$ can be contracted to nothing (i.e., to its boundary) and we can think of the unitary transformation as being the identity. As with the identity cobordism above, we can take such an identity transformation, glue the top to the bottom and obtain

$$Z(\Sigma \times S^1) = \mathrm{Dim}[V(\Sigma)] \tag{6.4}$$

I.e., the partition function Z is just the dimension of the Hilbert space of the wavefunction. This holds true even when Σ has marked points, or particles, in it.

6.2.1 Particles or No-Particles

In the same way that the ground state of a topological system "knows" about the types of anyons that can exist in the system, it is also the case that the TQFT in the absence of particles actually carries the same information as in the presence of particles. To see this consider a manifold $\mathcal{M}$ with labeled and directed world-lines L_i in them, as shown in Fig. 6.9. Now consider removing the world lines along with a hollow tube surrounding the paths that the world-lines follow as shown in the Figure. We now have a manifold with a solid torus removed for each world-line loop (think

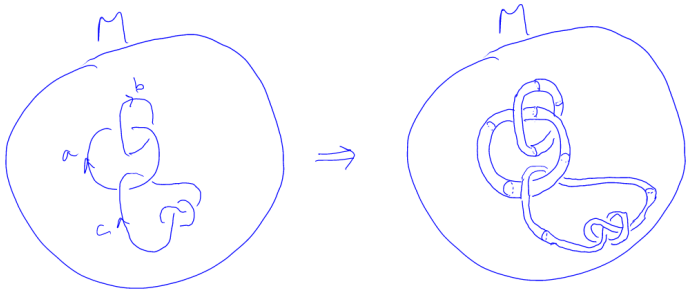


Figure 6.9: Removing the world-lines on the left along with a thickened tube. Imagine a worm burrowing along the path of the world lines and leaving a hollow hole.

of a big worm having eaten a path out of the manifold.) In this configuration, the boundary $\partial\mathcal{M}$ of the manifold $\mathcal{M}$ now contains the surface of these empty tubes — i.e, the surface of a torus T^2 for each world-line loop. Note that the empty tube is topologically a solid torus $D^2 \times S^1$ even if the world-line forms some knot. The statement that it forms a knot is some statement about the embedding of the S^1 loop in the manifold.

Note that the Hilbert space of the torus surface T^2 is in one-to-one correspondence with the particle types that can be put around the handle of the torus. Indeed, each possible state $|\psi_a\rangle$ of the torus surface corresponds to a picture like that of figure 6.10, where a particle of type a goes around the handle. Obviously, gluing such a torus back into the empty torus-shaped tube recovers the original picture of labeled world lines following these paths. We can think of this solid torus manifold as being a space-time history where $t = -\infty$ is the central core of the torus (the circle that traces the central line of the jelly filling of the donut) and the surface is the present time. Somewhere between $t = -\infty$ and now a particle of type a has been dragged around the handle.

The manifold with the tori excised from it (the right of fig. 6.9) contains all of the information necessary to give a partition function for the left of fig. 6.9 for *any* particle types that we choose to follow the given world lines. For the manifold on the right we have

$$Z(M) = \sum_{i,j,k} Z(M; i, j, k)\ \langle\psi_{L1,i}| \otimes \langle\psi_{L2,j}| \otimes \langle\psi_{L3,k}|$$

where $Z(M; i, j, k)$ is the partition function for the torus with three particle types i, j, k following the three world line loops L_1, L_2 and L_3, and the three wavefunctions are the corresponding boundary

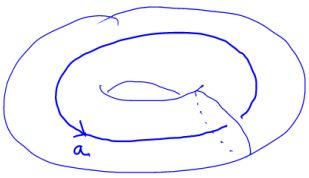


Figure 6.10: The possible wavefunctions $|\psi_a\rangle$ that we can have on the surface of the torus can be realized by having a world-line of a particle of type a going around the handle of the torus. We can call these $Z(\text{solid torus with } a \text{ running around handle}) = |\psi_a\rangle$

condition. Thus, if we want to extract $Z(a, b, c)$, where the particle lines are labeled with a, b, c we simply glue in the torus wavefunctions $|\psi_{L1,a}\rangle$, $|\psi_{L2,b}\rangle$, $|\psi_{L3,c}\rangle$.

6.3 Building 3-Manifolds

6.3.1 S^3 and the modular S-matrix

We will now consider building up 3-manifolds from pieces by gluing objects together. The simplest 3-manifold to assemble is the three sphere S^3. Remember that S^3 can be thought of as $\mathbb{R}^3$ compactified with a single point at infinity (the same way that S^2 is a plane, closed up at infinity — think of stereographic projection.) Recall also that a solid torus should be thought of as a disk crossed with a circle $D^2 \times S^1$. I claim that we can assemble S^3 from two solid tori

$$S^3 = (S^1 \times D^2) \cup (D^2 \times S^1)$$

There is a very elegant proof of this statement. Consider the 4-ball D^4. Topologically we have[2]

$$D^4 = D^2 \times D^2$$

Now applying the boundary operator ∂ and using the fact that the boundary operator obeys the Leibniz rule (i.e, it distributes like a derivative), we have

$$S^3 = \partial D^4 = \partial(D^2 \times D^2) = (S^1 \times D^2) \cup (D^2 \times S^1)$$

where we have used that the boundary of a disk is a circle, $\partial D^2 = S^1$. Note that the two solid toruses differ in that they have the opposite D^2 filled in. Note that the two solid tori here are glued together along a common $T^2 = S^1 \times S^1$ boundary. To see this note that

$$\partial(S^1 \times D^2) = S^1 \times S^1 = \partial(D^2 \times S^1)$$

The toruses are glued together meridian-to-longitude and longitude-to-meridian. (I.e., the contractable direction of one torus is glued to the non-contractable direction of the other, and vice versa) A sketch of how the two solid tori are assembled together to make S^3 is given in Fig. 6.11.

OK, so here we have two solid tori which are glued together on their boundaries to make up S^3, so we can write the partition function as the overlap between wavefunctions on the outside and and inside toruses.

$$Z(S^3) = \langle Z(D^2 \times S^1) | Z(S^1 \times D^2) \rangle$$

where the ψ's are the wavefunctions on the surface of the torus.

We can further consider including world lines around the non-contractable loops of the solid torus, as in Fig. 6.10. There is a different state on the surface of the torus for each particle type we

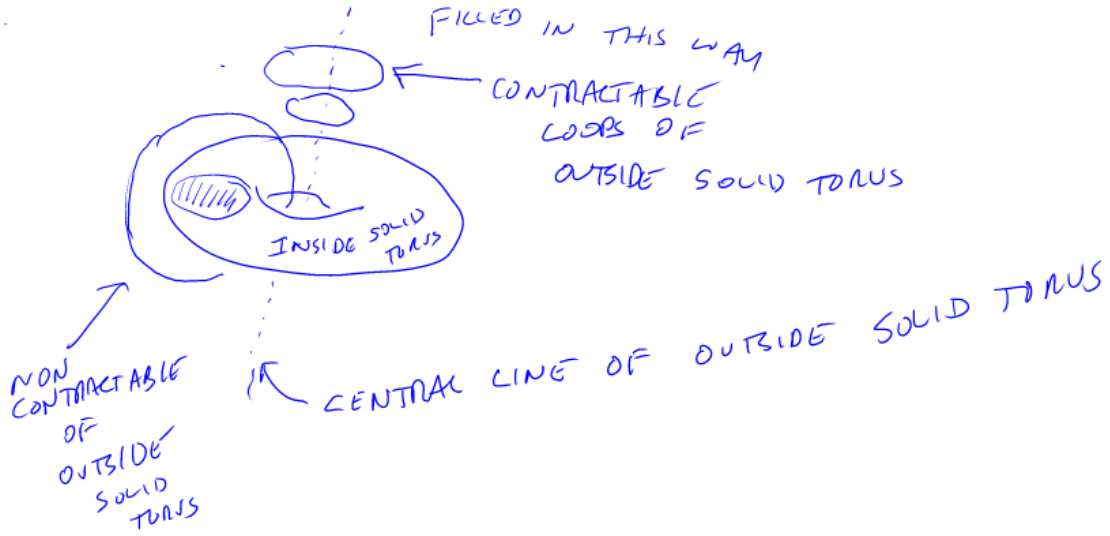


Figure 6.11: Assembling two solid toruses to make S^3. One torus (labeled "inside solid torus") is obvious. Here the contractable disk D^2 is shaded and the noncontractable S^1 is the handle. The remainder of space, including the point at infinity is the other "outside" solid torus. The central line of the handle (i.e., the noncontractable S^1 runs vertically through the center of the picture. It is a circle because it connects up at the point at infinity. All of the other non-contractable loops can be deformed continuously to the central line. The contractable disks (or contractable loops) are also drawn.

have running around the handle. We then assemble S^3 with these new solid tori and get an S^3 with two particle world lines linked together as shown in Fig. 6.12. Gluing the two tori together we get

$$Z(S^3; a \text{ loop linking } b \text{ loop}) = \langle Z(D^2 \times S^1; b) | Z(S^1 \times D^2; b) \rangle \equiv S_{ab} \tag{6.5}$$

This quantity S_{ab} is known as the **modular S-matrix**, and it is a very important quantity in topological theories as we shall soon see. Since this is an overlap we must have $0 \leq S_{00} \leq 1$ (by this we mean the element of the S-matrix where the vacuum particle is put around both handles — or no particle at all — we are using 0 to mean the vacuum now!). This tells us that

$$Z(S^3) = S_{00} \leq 1$$

And in fact, invoking unitarity, unless $S_{0a} = 0$ for all other particle types a (which is never true!) we must have S_{00} be strictly less than 1.

Note that the S-matrix is unitary, since it is simply giving a basis transformation between the two sets of wavefunction which both span the vector space $V(T^2)$ of the torus surface T^2 where the two solid toruses are glued together.

Soon we will construct a set of diagramatic rules to help us "evaluate" the matrix S_{ab}. These rules will be somewhat similar to the rules for the Kauffman invariant, only now we need to keep track of labels on world lines as well.

[2]Topologically it is easiest to think about D^n as being the interval $I = D^1$ raised to the n^{th} power. I.e., the disk, is topologically a filled-in square $D^2 = D^1 \times D^1$. The usual 3-ball is topologically a cube $D^3 = D^1 \times D^1 \times D^1$. The 4-ball is topologically a 4-cube $D^4 = D^1 \times D^1 \times D^1 \times D^1 = D^2 \times D^2$.

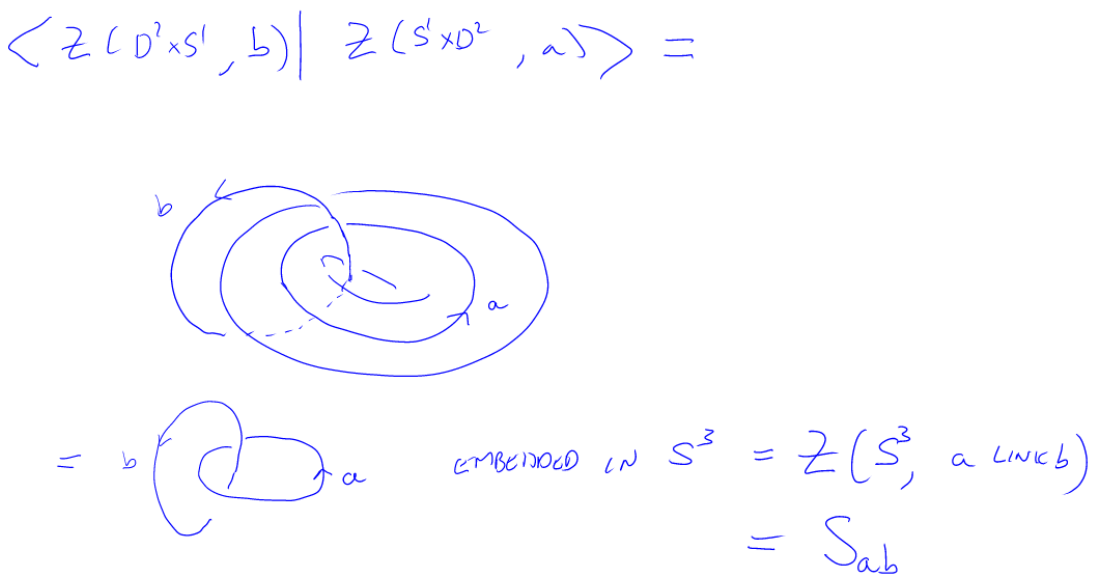


Figure 6.12: Here we assemble a partition function for S^3 with world lines of a linking b embedded in the S^3. To do this we glue together two solid tori each with a world line running around the handle. The end result is known as the modular S-matrix, and it gives a basis transform converting between the two bases which both span the hilbert space of the torus surface where the two solid tori are glued together.

6.3.2 $S^2 \times S^1$

There is another way we can put two solid tori together to make a closed manifold. Instead of attaching longitude-to-meridian and meridian-to-longitude, we instead attach meridian-to-meridian and longitude-to-longitude. (This is perhaps even simpler!) See Figure 6.13 Here we want to show

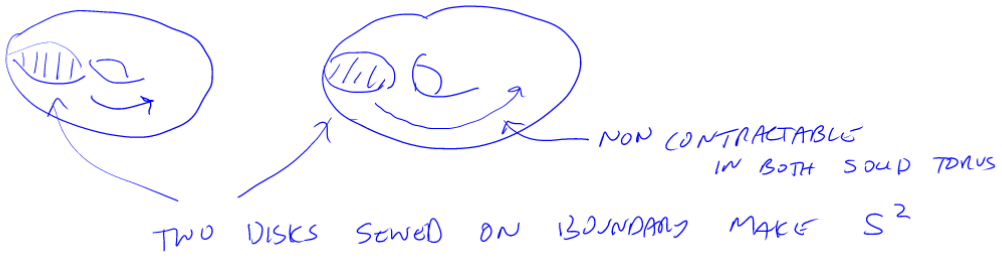


Figure 6.13: Assembling two solid tori to make $S^2 \times S^1$. Here the two contractable disks D^2 are sewed together along their boundaries to make S^2.

that

$$S^2 \times S^1 = (D^2 \times S^1) \cup (D^2 \times S^1)$$

The sewing together is again done along the common boundary $T^2 = S^1 \times S^1$. The S^1 factors in both solid tori are the same, and both of the D^2 have the same S^1 boundary. Thus we are sewing togther two disks D^2 along their S^1 boundaries to make a 2-sphere S^2 (imagine cutting a sphere in half and getting two disks). As in the previous case, we can put world lines through the handles of the toruses if we want.

In this case we have

$$\langle Z(D^2 \times S^1; b) | Z(D^2 \times S^1; a) \rangle = \delta_{ab}$$

The reason it is a delta function is that both the bra and ket are really the same wavefunctions (we have not switched longitude to meridian). So except for the conjugation we should expect that we are getting the same basis of states for both toruses

In particular, we have the case where we put no particle (the vacuum) around both handles, we have (i.e., $a = b = I$)

$$\langle Z(D^2 \times S^1) | Z(D^2 \times S^1) \rangle = \delta_{ab} = 1$$

So we have the result

$$Z(S^2 \times S^1) = 1$$

Note that this agrees with our prior statement Eq. 6.4 that Z for any manifold $\times S^1$ should be the dimension of the Hilbert space for that manifold along with Eq. 6.3 that the dimension of the Hilbert space on a sphere is 1.

6.4 Surgery and More Complicated Manifolds

The understanding of 3-manifolds is a very difficult problem[3] The methods that I will describe here are fairly powerful for describing 3-manifold topology (as well as 4-manifold topology). For more detailed discussion of Surgery and Kirby Calculus, see the book by Kirby or the book by Gompf and Stipcitz.

In order to describe complicated manifolds it is useful to think in terms of so-called surgery. Similar to what we were just discussing — assembling manifold by gluing pieces together – the idea of surgery is that we remove a part of a manifold and we glue back in something different. (imagine replacing someone's foot with a hand! Prehensile toes could be useful I suppose!)

The general scheme of surgery is to first write a manifold as the union of two manifolds with boundary sewed along their common boundaries. First let us split a closed manifold $\mathcal{M}$ into two pieces with boundary, $\mathcal{M}_1$ and $\mathcal{M}_2$ such that they are sewed together along their common boundary $\partial\mathcal{M}_1 = \partial\mathcal{M}_2^*$. So we have

$$\mathcal{M} = \mathcal{M}_1 \cup_{\partial\mathcal{M}_1} \mathcal{M}_2$$

We then find another manifold with boundary $\mathcal{M}_2'$ whose boundary matches $\mathcal{M}_2$, i.e,

$$\partial\mathcal{M}_2 = \partial\mathcal{M}_2'$$

We can then replace $\mathcal{M}_2$ with $\mathcal{M}_2'$. To construct a new closed manifold $\mathcal{M}'$ as

$$\mathcal{M}' = \mathcal{M}_1 \cup_{\partial\mathcal{M}_1} \mathcal{M}_2'$$

We say that we have performed surgery on $\mathcal{M}$ to obtain $\mathcal{M}'$.

6.4.1 Simple example of surgery on a 2-manifold

To give an example of this consider the sphere $\mathcal{M} = S^2$ as shown in Fig. 6.14. Here we write the sphere as the union of two disks $\mathcal{M}_2 = D^2 \cup D^2$ and the remainder of the sphere $\mathcal{M}_1 = S^2 - (D^2 \cup D^2)$. These are glued along their common boundary $S^1 \cup S^1$.

Now we ask the question of what other 2-manifolds have the same boundary $S^1 \cup S^1$. There is a very obvious one, the cylinder surface! Let us choose the cylinder surface $\mathcal{M}_2' = S^1 \times I$ where I is the interval (or D^1). It has boundary $\partial\mathcal{M}_2' = S^1 \cup S^1$. Thus we can glue it in place where we remove $\mathcal{M}_2$,as shown in Fig. 6.15. The resulting manifold $\mathcal{M}'$ is the torus T^2

$$T^2 = \left(S^2 - (D^2 \cup D^2)\right) \cup_{S^1 \times S^1} (S^1 \times I)$$

[3]Many important results have been discovered recently. Particularly Perelman's proof of Poincare Conjecture, along with the methods he used are apparently extremely powerful. But I can't say I understand these!.

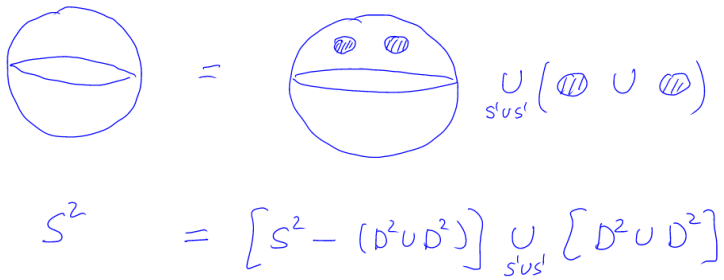


Figure 6.14: Writing a sphere $\mathcal{M} = S^2$ as the union of two manifolds glued along their boundaries. $\mathcal{M}_2$ is the union of two disks $D^2 \cup D^2$. $\mathcal{M}_1 = S^2 - (D^2 \cup D^2)$ is the remainder. The two manifolds are glued along their common boundary $S^1 \cup S^1$.

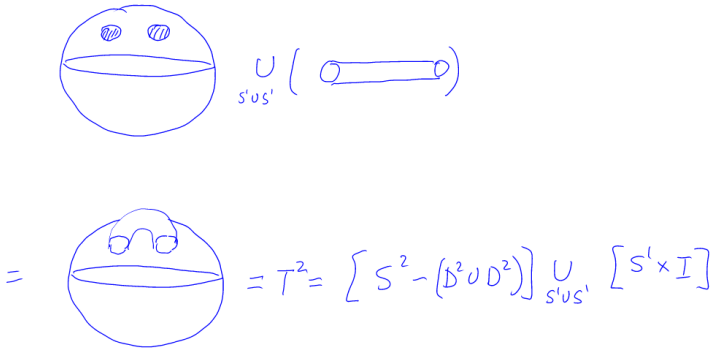


Figure 6.15: Gluing the cylinder surface $\mathcal{M}_2' = S^1 \times I$ to the manifold $\mathcal{M}_1 = S^2 - (D^2 \cup D^2)$ along their common boundary $S^1 \cup S^1$ gives the torus T^2.

Thus we have surgered a sphere and turned it into a torus. Note that there is another way to think of this procedure. If $\mathcal{M} = \partial\mathcal{N}$ then surgery on $\mathcal{M}$ is the same as attaching a handle to $\mathcal{N}$. In the case we just considered we would take $\mathcal{N} = D^3$ the 3-ball (sometimes denoted B^3). And we attach a handle $D^2 \times I$, the solid cylinder. We obtain the new manifold $\mathcal{N}'$ which is the solid torus, whose boundary is T^2 the torus surface. This is written out in the diagram Fig. 6.16

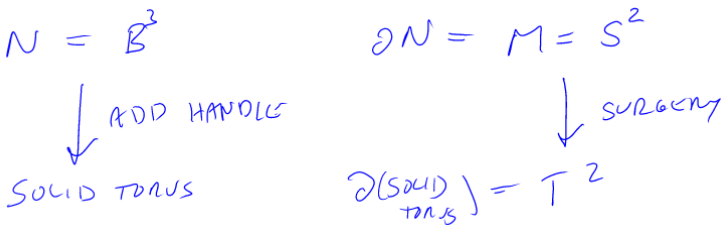


Figure 6.16: Surgery on a manifold $\mathcal{M} = \partial\mathcal{N}$ is the same as handle attaching on the manifod $\mathcal{N}$.

6.4.2 Surgery on 3-manifolds

OK, here is the part that is guaranteed to make your head explode.

Start with a 3-manifold $\mathcal{M}$, such as perhaps the $\mathbb{R}^3$ space around us, or maybe S^3. Now consider a solid torus

$$\mathcal{M}_2 = D^2 \times S^1$$

embedded in this manifold. The surface $\partial\mathcal{M}_2 = S^1 \times S^1 = T^2$ is a torus surface. Now, there is another torus with exactly the same surface. It is

$$\mathcal{M}_2' = S^1 \times D^2$$

These two solid tori differ in that they have opposite circles filled in. Both have the same $S^1 \times S^1$ surface, but $\mathcal{M}_2$ has the first S_1 filled in whereas $\mathcal{M}_2'$ has the second S_1 filled in.

The idea of surgery is to remove $\mathcal{M}_2$ and replace it with $\mathcal{M}_2'$ to generate a new manifold $\mathcal{M}'$ with no boundary. (Stop here, think about what we have done. Collect pieces of your exploded head) The reason your head should explode is because it is hard to visualize the end result because the new structure is not embeddable within the original $\mathbb{R}^3$. This is torus surgery on a 3-manifold, and it is called Dehn surgery. Another way to describe what we have done is that we have removed a torus, switched the meridian and longitude (switched the filled-contractable and the unfilled-uncontractable) and then glued it back in. In fact, one can make more complicated transformations on the torus before gluing it back in (and it is still called Dehn surgery) but we will not need this.

It is worth noting that the torus we removed could be embedded in a very complicated way within the original manifold — i.e, it could follow a complicated, even knotted, path, as in the figure on the right of Fig. 6.9.

Lickorish-Wallace Theorem

This is quite an important theorem of topology[4].

Theorem: Starting with S^3 one can obtain any closed 3-manifold by performing successive torus surgeries, where the initial torus may be nontrivially embedded in the manifold (i.e, it may folllow some knotted path)

So one has the following procedure. We start with a link (some knot possibly of several strands), embedded in S^3. Thicken each line to a torus. Excise each of these toruses, and replace them by a torus with longitude and meridian switched. If you start with the correct link, you can get any possible 3-manifold. We summarize with

$$\text{Link in } S^3 \quad \overset{\text{surger}}{\longrightarrow} \quad \text{Some } M^3$$

Kirby Calculus

It is not the case that all topologically different links, when surgered, give topologically different manifolds. Fortunately, the rules for which knots give the same manifolds have been worked out. These rules, known as Kirby calculus, are stated as a set of transformations on a link which change the link, but leave the resulting manifold unchanged. There are several different sets of moves that can be taken as "elementary" moves, but perhaps the simplest elementary moves are known as Kirby moves:

(1) **Blow up/ Blow Down** One can add or remove an unlinked loop with a single twist, as shown in Fig. 6.17.

(2) **Handle-Slide** A string can be broken open and pulled along the full path of another string, and then reconnected. See Fig. 6.18.

Two links describe the same manifold if and only if one link can be turned into the other by a sequence of these Kirby moves.

[4]In Witten's groundbreaking paper, he states the theorem without citation and just says "It is a not too deep result..". Ha!

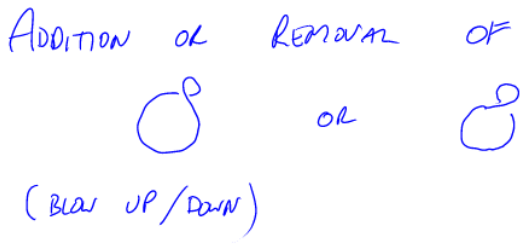


Figure 6.17: Blow up/ Blow down.

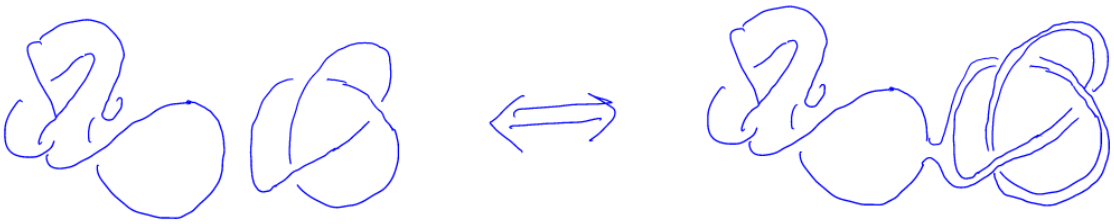


Figure 6.18: A handle-slide move

Witten-Reshitikhin-Turaev Invariant

Given knowledge of the rules of Kirby calculus, in order to construct a manifold invariant for three manifolds, one need only construct a knot invariant that is invariant under Kirby moves. Being that the Chern-Simons path integral is not really well defined as a path integral, it turns out that this scheme is a way to make mathematically rigorous the manifold invariants of Chern-Simons theory.

Without ever saying the words "path integral" or "Chern-Simons action" we can think of an anyon theory as simply a way to turn a link of labeled world lines into a number. (Like evaluating a knot invariant, but with rules for labeled links). It turns out that for any well behaved anyon theory one can put together a combination of world-line types that will obey the Kirby calculus and therefore allow one to construct a manifold invariant.

The first Kirby move (The blow up/blow down) does not sound so hard to finagle just by using some normalization factor for each twist and loop. The second Kirby move seems harder to achieve, but can be achieved if one uses the so-called Kirby color combination (Or Ω string)

$$|\Omega\rangle = \sum_a S_{0a}|a\rangle$$

where here we mean that we are summing over particle types a, and S is the modular S-matrix. Diagrammatically we have Fig. 6.19 It turns out (and we will show this later) that the corresponding knot invariant that comes from evaluating a knot of Kirby color is invariant under handle-slides. The manifold invariant that results from evaluating the corresponding knot invariant of the Kirby-color string is known as the Witten-Reshitikhin-Turaev invariant and it gives a rigorous re-definition of the Chern-Simons manifold invariants defined by Witten.

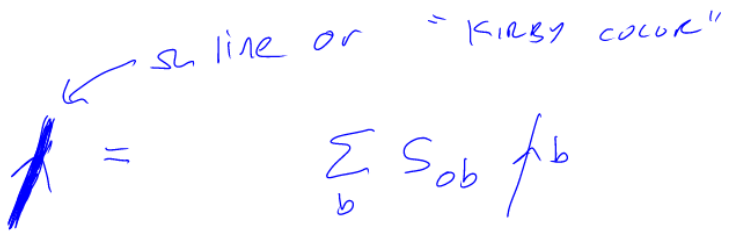


Figure 6.19: A String of Kirby Color is a weighted superposition of all anyon string types

Chapter 7

Fusion and Structure of Hilbert Space

So far we have discussed that each 2-d surface (a slice of a 3-d space-time manifold) has associated with it a Hilbert space. In the case where there are particles in this surface, the dimension of the Hilbert space will reflect the nature of the particles. We now seek to understand the structure of the this Hilbert space and how it depends on the particles.

7.1 Basics of Particles and Fusion — the abelian case

Particle types: There should be a finite set of labels which we call particle types. For now, let us call them a, b, c, etc.

Fusion: World lines can merge to give fusion diagrams, or do the reverse, which we call a splitting. If an a particle merges with b to give c, we write $a \times b = b \times a = c$. This is shown diagrammatically in Fig. 7.1.

Figure 7.1: Left: Fusion of two particle types to make a third $a \times b = c$. Right: Splitting of a particle type to make two others.

It should be noted that we can think of two particles as fusing together even if they are not close together. We need only draw a circle around both particles and think about the "total" particle type inside the circle. For example, we sometimes draw pictures like shown in Fig. 7.2.

Figure 7.2: Fusion of two particle types to make a third $a \times b = c$. The two particles need not be close to each other.

For example, in our abelian anyon model of charges and fluxes, if the statistical angle was $\theta = \pi p/m$ (p and m relatively prime) we had m species $a = (aq, a\Phi)$ for $a = 0 \ldots m-1$, where $q\Phi = \pi p/m$. The fusion rules were simply addition modulo m. That is $a \times b = (a+b) \mod m$.

Identity: Exactly One of the particles should be called the identity or vacuum. This is written as 1 or 0 or I or e. It fuses trivially with any particles.

$$a \times I = a$$

for all possible a. In the above abelian model we should think of the identity as being no charge and no flux. Fusion with the identity is depicted schematically in fig. 7.3

Figure 7.3: Two depictions of fusion of a particle with the identity $a \times I = a$. On the right, the empty space with a light dotted circle is supposed to indicate the identity. The circle surrounding a and the identity, has particle type a.

Antiparticles: Each particle a should have a unique antiparticle which we denote as $\bar{a}$. The antiparticle is defined by $a \times \bar{a} = I$. A particle going forward should be equivalent to an antiparticle going backwards as shown in Fig. 7.4.

Figure 7.4: A particle going forward should be equivalent to an antiparticle going backwards

Fusion to the identity can be thought of as particle turning around as shown in Fig. 7.5.

Figure 7.5: A particle going forward should be equivalent to an antiparticle going backwards

A particle may be its own inverse, in which case we do not need to draw arrows on its world lines. An example of this in our charge-flux model would be the $a = 2$ particle with $m = 4$ for $\theta = \pi p/m$.

7.2 Multiple Fusion Channels - the Nonabelian Case

For the nonabelian theories as we have discussed above, the dimension of the Hilbert space must increase with the number of particles present. How does this occur? In nonabelian models we have multiple possible fusion channels

$$a \times b = c + d + \ldots$$

meaning that a and b can come together to form either a c or a d or .... See Fig. 7.6. A theory is nonabelian if *any* two particles fuse in such a way that there are multiple possible fusion channels (i.e., there is more than one particle listed on the right hand side).

Figure 7.6: Multiple possible fusion channels

If there are s possible fusion channels for $a \times b$, then the two particles a and b have an s dimensional Hilbert space (part of what we called $V(\Sigma)$).

So what is this Hilbert space. A slightly imperfect analogy is that of angular momentum addition. We know the rule for adding spin 1/2

$$\frac{1}{2} \otimes \frac{1}{2} = 0 \oplus 1$$

which tells us that two spin 1/2's can *fuse* to form a singlet or a triplet. As with the case of spins, we can think about the two particles being in a wavefunction such that they fuse in one particular fusion channel or the other — even if the two particles are not close together. The singlet or $J = 0$ state is the identity here, it has no spin at all. The analogy spins is not exact though — unlike the case of spins, the individual particles have no internal degrees of freedom (analogous to the 2-states of the spin 1/2), nor do any results of fusion have an m_z degree of freedom (like a triplet would).

Locality

The principle of locality is an over-arching theme of anyon physics (if not of physics altogether).

The quantum number (or "charge") of a particle is locally conserved in space. So for example in Fig. 7.7, on the left, a particle a is propagating along and suddenly something complicated happens locally. If only a single particle comes out of this region it must also be a particle of type a. (If two particles come out of this region, we could have a split into two other species as in the right of 7.1). This principle we can call the **no transmutation** principle. It allows us to conclude that the complicated picture on the left of Fig. 7.7 must be equal to some constant times the simple propagation of an a particle as shown on the right.

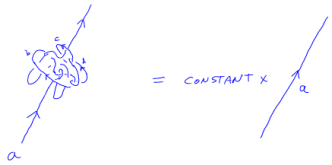


Figure 7.7: Multiple possible fusion channels

If two particles (maybe far away from each other) fuse to some other particle type (in a case where multiple fusion channels are available) it is not possible to determine what the fusion channel is by measuring only one of the initial particles. In order to determine the fusion channel of the two particles, you have to (for example) perform an interference measurement that surrounds both of these particles. The fusion channel is *local* to the pair.

Similarly, if we have some particles, b and c and they fuse to d (See Fig. 7.8), no amount of braiding b around c will change this overall fusion channel. Similarly, if these two then fuse with a to give an overall fusion channel f, no amount of braiding a, b and c will change the overall fusion channel f. However, if a braids with b and c, then the fusion of b and c might change, subject to the constraint that the overall channel of all three particles remains f.

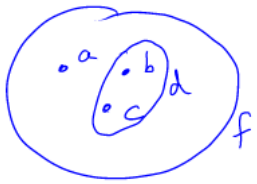


Figure 7.8: In this picture b and c fuse to d. Then this d fuses with a to give an overall fusion channel of f.

Antiparticles with Multiple Fusion Channels: When we have multiple fusion channels we define antiparticles via the principle that a particle *can* fuse with it antiparticle to give the identity, although other fusion channels may be possible.

$$a \times \bar{a} = I + \text{ Other Fusion Channels}$$

It should be the case that for each particle a there is a unique $\bar{a}$ that can fuse to the identity.

7.2.1 Example: Fibonacci Anyons

Perhaps the simplest nonabelian example is the anyon system known as Fibonacci[1] Anyons. Something very close to this is thought to occur in the so-called $\nu = 12/5$ quantum Hall state. It is closely related to the $SU(2)_3$ Chern-Simons theory[2]

In this system the particle set includes only two particles, the identity I and a nontrivial particle which is often called τ.

$$\text{Particle types} = \{I, \tau\}$$

The fusion rules are

$$I \times I = I$$
$$I \times \tau = \tau$$
$$\tau \times \tau = I + \tau$$

The first two of these rules hardly need to be written down (they are implied by the required properties of the identity). It is the final rule that is very nontrivial. Note that it implies that τ is its own antiparticle $\tau = \bar{\tau}$ which means we do not need to put arrows on world lines.

With two Fibonacci anyons the Hilbert space is two dimensional (since the two particles can fuse to I or τ. See Fig. 7.9.

[1] Fibonacci's real name was Leonardo Bonacci and he lived around 1200 AD.

[2] Fibonacci anyons can be described exactly by the $G2$ level 1 Chern-Simons theory. This involves a messy Lie algebra called $G2$. The $SU(2)_3$ Chern-Simons theory contains some additional particles besides the Fibonacci particles, but ignoring these, it is the same.

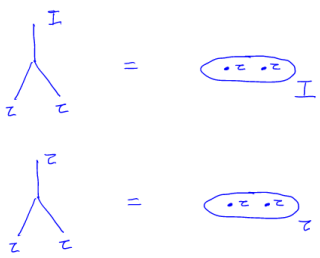


Figure 7.9: Two different notations for the two different fusion channels of two Fibonacci anyons

With three fibonacci anyons the Hilbert space is 3 dimensional, as shown in Fig. 7.10. The key thing to notice is that if the first two particles fuse to τ, then this combination acts as being a single particle of overall charge τ — it can fuse with the third τ in two ways.

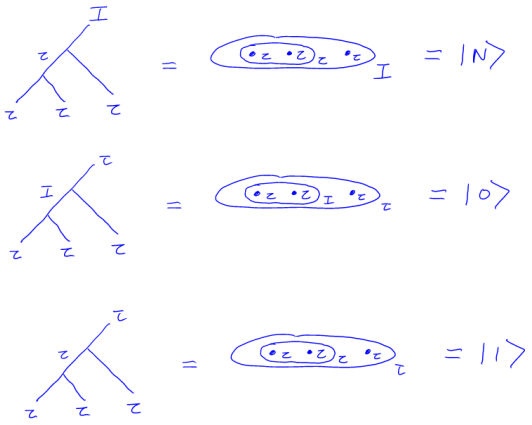


Figure 7.10: Notations for the three different fusion channels of three Fibonacci anyons. The notation $|N\rangle, |0\rangle$ and $|1\rangle$ are common notations for those interested in topological quantum computing with Fibonacci anyons!

There are two states in the Hilbert space of three anyons (labeled $|0\rangle$ and $|1\rangle$ in Fig. 7.10) which both have an overall fusion channel of τ. As mentioned above, due to locality, no amount of braiding amongst the three particles will change this overall fusion channel. Further since in these two basis states the first two particles furthest left are in an eigenstate, either I in state $|0\rangle$ or τ in state $|1\rangle$) no amount of braiding of the first two particles will change that eigenstate. However, as we will see below, if we braid the second particle with the third, we can then change the quantum number of the first two particles and rotate between $|0\rangle$ and $|1\rangle$.

For our Fibonacci system, with 2 particles the Hilbert space is 2 dimensional. With 3 particles the Hilbert space is 3 dimensional. It is an easy exercise to see that with 4 particles the Hilbert space is 5 dimensional, and with 5 particles, 8 dimensional and so forth. This pattern continues following the Fibonacci sequence (Try to show this!).

Since the N^{th} element of the Fibonacci sequence for large N is approximately

$$\text{Dim of } N \text{ Anyons} = \text{Fib}_N \sim \left(\frac{1+\sqrt{5}}{2}\right)^N$$

we have the quantum dimension of this particle is $d = (1+\sqrt{5})/2$.

7.2.2 Example: Ising Anyons

The Ising anyon system is extremely closely related to $SU(2)_2$ Chern-Simons theory[3], and this general class of anyon is believed to be realized in the $\nu = 5/2$ quantum Hall state, topological superconductors, and other so-called Majorana systems.

The Ising theory has three particle types.

$$\text{Particle types} = \{I, \sigma, \psi\}$$

The nontrivial fusion rules are

$$\begin{aligned} \psi \times \psi &= I \\ \psi \times \sigma &= \sigma \\ \sigma \times \sigma &= I + \psi \end{aligned}$$

where we have not written the outcome of any fusion with the identity, since the outcome is obvious. Again, each particle is its own antiparticle $\psi = \bar{\psi}$ and $\sigma = \bar{\sigma}$ so we need not put arrows on any world-lines.

Fusion of many ψ particles is fairly trivial, since each pair fuses to the identity in only one way (we say that ψ is an abelian particle, although the theory is nonabelian).

Fusion of many σ particles is nontrivial . The first two σs can either fuse to I or ψ, but then when the third is included the overall fusion channel must be σ (since fusing σ with either ψ or I gives σ. Then adding a fourth σ to this cluster whose overall quantum number is σ again gives two possible outcomes. See the fusion tree in Fig.7.11

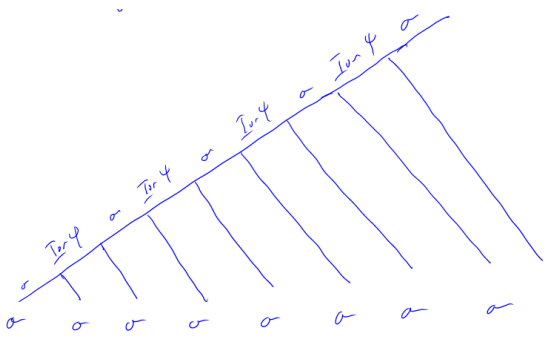


Figure 7.11: The Ising Fusion Tree.

The total number of different fusion channels for N σ-particles is $2^{N/2}$. To see this in another way, we can group σ particles together in pairs where each pair gives either ψ or I, so two σ particles comprises a qubit. Then the I's and ψ's fuse together in a unique way. Since the Hilbert space dimension is $(\sqrt{2})^N$ so the quantum dimension of the σ particle is $d = \sqrt{2}$.

7.3 Fusion and the N matrices

We are well on our way to fully defining an anyon theory. A theory must have a finite set of particles, including a unique identity I, and each particle having a unique inverse.

The general fusion rules can be written as

$$a \times b = \sum_c N_{ab}^c \; c$$

[3]The fusion rules are the same, but there are some spin factors which differ.

where the N's are known the fusion multiplicities. N_{ab}^c is zero if a and b cannot fuse to c. N_{ab}^c is one if we have $a \times b = \ldots + c + \ldots$, and c only occurs once on the right hand side. If c occurs more than once on the right hand side, then N_{ab}^c simply counts the number of times it occurs.

What does it mean that a particle type can occur more than once in the list of fusion outcomes? It simply means that the fusion result can occur in multiple orthogonal ways .[4] in which case a diagram with a vertex showing a and b fusing to c should also contain an index at the vertex indicating which of the possible c fusion channels occurs. For most simply anyon theories N_{ab}^c is either 0 or 1, and indeed, we will not consider the more complicated case in any examples below.

In terms of these fusion multiplicity matrices we have

$$N_{ab}^c = N_{ba}^c$$

which is just commutativity of fusion $a \times b = b \times a$.

Fusion with the identity implies

$$N_{aI}^b = \delta_{ab}$$

Uniqueness of inverse implies that for each a there is exaclty one anyon b such that

$$N_{ab}^I = 1$$

and we call this b the inverse of a and denote it $\bar{a}$. Note that we can never have $N_{ab}^I > 1$. Also note that a can be the same as $\bar{a}$.

If we are to fuse, say, five particles a together, we can do so via a tree as shown in Fig. 7.12.

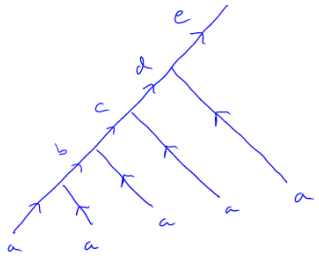


Figure 7.12: Fusing five a particles together

To find the dimension of the Hilbert space, we write

$$\text{Dimension of fusing five } a \text{ particles} = \sum_{bcde} N_{aa}^b N_{ba}^c N_{ca}^d N_{da}^e$$

and we identify each factor of N as being one of the vertices in the figure.

We recall that the quantum dimension d_a of the particle a is defined via the fact that the Hilbert space dimension should go as d_a^N where N is the number of a particles fused together. To find this we should think of N_{ab}^c as a matrix N_a with indices b and c, i.e, we write $[N_a]_b^c$. We then have that

$$d_a = \text{largest eigenvalue of } [N_a]$$

Example of Fibonacci: The fusion matrix for the τ particle in the Fibonacci theory is

$$N_\tau = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

[4] While this does not occur for angular momentum addition of $SU(2)$ (and also will not occur in Chern-Simons theory $SU(2)_k$ correspondingly) it is well known among high energy theorists who consider the combination of representations of $SU(3)$. Recall that

$$8 \otimes 8 = 1 \oplus 8 \oplus 8 \oplus 10 \oplus \bar{10} \oplus 27$$

and the 8 occurs twice on the right.

where here the first row and first column represent the identity and the second row and second column represent τ. The first row of this matrix says that fusing τ with identity gives back τ and the second row says that fusing τ with τ gives I and τ. It is an easy exercise to check that the largest eigenvalue of this matrix is indeed $d_\tau = (1+\sqrt{5})/2$.

Example of Ising: The fusion matrix for the σ particle in the Fibonacci theory is

$$N_\sigma = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

where the first row and column represent the identity, the second row and column represent σ and the third row and column represent ψ. So, for example, the second row here indicates that $\sigma \times \sigma = I + \psi$. Again, it is an easy exercise to check that the largest eigenvalue of this matrix is $d_\sigma = \sqrt{2}$.

Associativity: It should be noted that the fusion multiplicity matrices N are pretty special since the outcome of a fusion should not depend on the order of fusion. I.e., $(a \times b) \times c = a \times (b \times c)$.

For example, let us try to calculate how many ways $a \times b \times c$ can give an outcome of e. We can either try fusing $a \times b$ first as on the left of Fig. 7.13 or we can try fusing b and c first as on the right. Correspondingly to these two possibilities we have

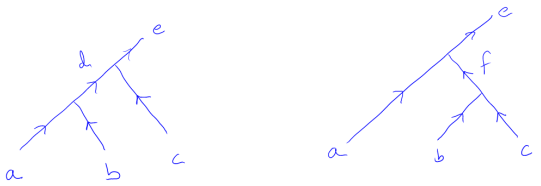


Figure 7.13: Fusing $(a \times b) \times c$ should be equivalent to $a \times (b \times c)$

$$\sum_d N_{ab}^d N_{cd}^e = \sum_f N_{cb}^f N_{af}^e$$

Again, thinking of N_{ab}^c as a matrix labeled N_a with indices b and c, this tells us that

$$[N_a, N_c] = 0$$

So all of the N matrices commute with each other. This means they are all simultaneously diagonalizable,

$$[U N_a U^{-1}]_{xy} = \delta_{xy} \lambda_x^{(a)} \tag{7.1}$$

and all N_a's get diagonalized with the same U. Surprisingly (as we will see below!) the matrix U is precisely the modular S-matrix we discussed above!

7.4 Fusion and Hilbert Space

The structure of fusion rules can be used to calculate the ground state degeneracy of wavefunctions on certain 2-dimensional manifolds.

Let us start by considering the sphere S^2, and let us assume that there are no anyons on the surface of the sphere. As mentioned previously there is a unique ground state in this situation because there are no non-contractable loops. The dimension of the Hilbert space is just 1

$$\text{Dim } V(S^2) = 1$$

This will be the starting point for our understanding. We should assume that all other configurations (change of topology, adding particles etc) should be related back to this reference configuration.

Now let us consider the possibility of having a single anyon on the sphere. In fact such a thing is not possible because you can only pair create particle-antiparticle pairs (the total anyon charge must be conserved — i.e., everything on the sphere must fuse together to total quantum number of the identity). Thus, we have

$$\text{Dim } V(S^2 \text{ with one anyon}) = 0$$

Another way to explain this is to realize that, since particle-antipartiles are made in pairs, there is no space-time history that could prepare the state with just a single particle on the sphere!

We can however consider the possibility of two anyons on a sphere. We can create an a and $\bar{a}$ particles, and since these two particles must fuse back to the identity in a unique way we have

$$\text{Dim } V(S^2 \text{ with one } a \text{ and one } \bar{a}) = 1$$

The two particles must be antiparticles of each other, otherwise no state is allowed and the dimension of the Hilbert space is zero. This is a general principle, we must require that the fusion of all the particles on the sphere must be the vacuum, since these particles must be (at some point) pulled from the vacuum.

Now we could also imagine puncturing the sphere to make a hole where the particles were. (Note that the twice punctured disk, without labeling the particle type looks like punctured disc in Fig.6.14.). In the spirit of what we did in section 6.2.1 we could re-fill the hole with any particle type. However, if we refill one hole with a particular particle type a, then the other hole must get only filled in with the anti-particle type $\bar{a}$. Nonetheless, we can conclude that

$$\text{Dim } V(S^2 \text{ with two unlabled punctures}) = \text{Number of particle types}$$

Now consider the procedure shown in Fig. 7.14. We start with the twice punctured sphere. The two punctures can be labeled with any particle-antiparticle pair labels. We can then deform the sphere to sew the two punctures together in a procedure that is identical to the surgery described in Fig. 6.15. The result of this surgery is to give the torus surface T^2 and we conclude that we should have

Figure 7.14: Surgering the twice punctured sphere into a torus. This is the gluing axiom in action.

$$Z(T^2 \times S^1) = \text{Dim } V(T^2) = \text{Number of particle types}$$

as we have already discussed. The general rule of surgery is that two punctures can be sewed together when they have opposing particle types.

This is exactly the gluing axiom of the TQFT. Although we are doing this surgery on a 2-dimensional surface, we should realize that there is also a time direcction, which we have implicitly assumed is compactified into to S^1. Thus we are sewing together the 2-surface (puncture$\times S^1$) with another 2-surface (puncture$\times S^1$), and the inner product between the two wavefunctions on these two-surfaces assures that the quantum number on these two punctures are conjugate to each other.

We can continue on to consider a sphere with three particles. Similarly we should expect that the three particle should fuse to the identity as shown in Fig. 7.15

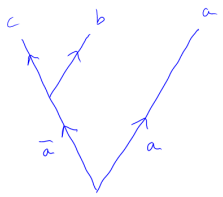


Figure 7.15: Three particles that fuse to the identity

We can then think of the sphere with three particles similarly as being a sphere with three labeled punctures which is known as a "pants", for reasons that are obvious in Figure. 7.16. It turns out that any orientable 2-dimensional manifold (except S^2 or T^2 which we have already considered can be constructed by sewing together the punctures of pants diagrams. For example, in Fig. 7.17 we sew together to pants to obtain a two handled torus.

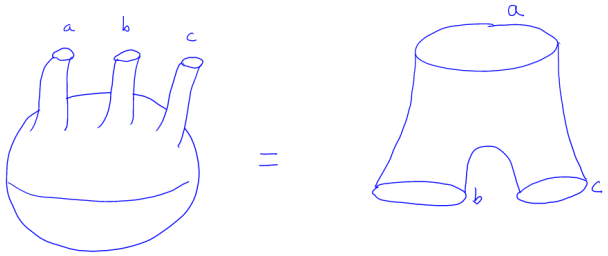


Figure 7.16: A three-times punctured sphere is known as a "pants" diagram.

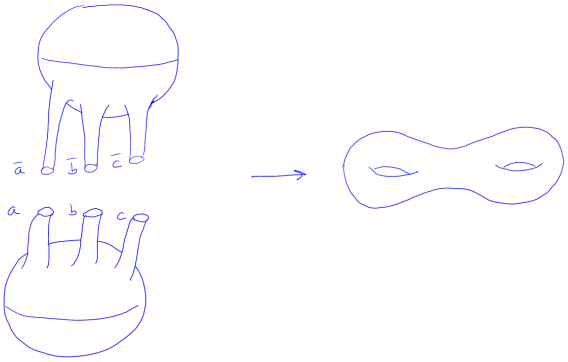


Figure 7.17: A three-times punctured sphere is known as a "pants" diagram.

To find the ground state degeneracy of the two handled torus Dim V(Two handled Torus) = Z(Two handled Torus $\times S^1$), we simply need to figure out the number of possible fusion channels where we could satisfy $a \times b \times c \to I$ and $\bar{a} \times \bar{b} \times \bar{c} \to I$ – this is precisely $N_{ab}^{\bar{c}}$ (which is also the same as $N_{\bar{a}\bar{b}}^{c}$). Essentially we are just looking at the number of ways we can assign labels to the punctures when we glue the objects together[5]. Looking at the fusion diagram 7.15 we then have

$$\text{Dim}V(\text{Two handled Torus}) = \sum_{abc} N_{bc}^{\bar{a}} N_{\bar{b}\bar{c}}^{a}$$

[5]If $N_{bc}^{\bar{a}} = 2$ then we need to count this configuration twice!

Example: Fibonacci Anyons on The Two-Handled Torus

With the Fibonacci fusion rules, there the five ways we can fuse three particles and get the identity.

$$I \times I \times I \to I$$
$$I \times \tau \times \tau \to I$$
$$\tau \times I \times \tau \to I$$
$$\tau \times \tau \times I \to I$$
$$\tau \times \tau \times \tau \to I$$

Note: Here there are several things to note about the notation. Here, the order of the three elements being multiplied together does matter, since the first element represents the first hole etc. Secondly, I use the $\to$ notation to indicate that it is a particular fusion channel out of many that could be possible. For example, if $\tau \times \tau = I + \tau$ to indicate that we mean the two τ's to fuse in the I channel only, I will write $\tau \times \tau \to \tau$.

At any rate, there are 5 possible labelings of the punctures that allow overall fusion to the identity. These are matched together on both sides of the diagram of 7.17 and we conclude that in the Fibonacci theory we have

$$Z(\text{Two Handled Torus} \times S^1) = \text{Dim}V(\text{Two handled Torus}) = 5$$

7.5 Change of Basis and F-symbols

As mentioned in Fig. 7.13, one can describe the same space in two different ways. If we are considering the space spanned by the fusion of $a \times b \times c$ as in the figure, we can describe the space by how a fuses with b (the value of d on the left of the figure), or by how b fuses with c (the value of f in the figure). Either of these two descriptions should be able to describe the space, but in different bases. We define the change of basis as a set of matrices called F. See figure 7.18

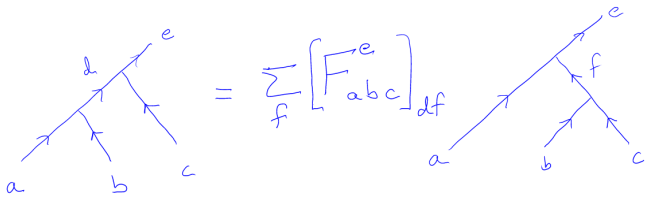


Figure 7.18: The F matrix makes a change of basis between the two differnt ways of describing the space spanned by the fusion of three anyon charges a, b, and c when they all fuse to e.

This idea of change of basis is familiar from angular momentum addition — where the F-matrix is known as a $6j$ symbol (note it has 6 indices). One can combine three objects with L^2 angular momenta values a, b and c in order to get L^2 angular momentum e, and quite similarly you can descibe this space in terms of a combined with b to get d or in terms of b combined with c to get f. (In fact, even when studying TQFTs, sometimes people refer to F-matrices as $6j$ symbols.)

Example: Fibonacci Anyons

Again we turn to the example of Fibonacci anyons for clarification of how this works. We imagine fusing together three τ particles. As shown in Fig. 7.10, there is a single state $|N\rangle$ in which the three fuse to the identity I. It should not matter if we choose to fuse the leftmost two anyons first, or the rightmost two. In either case there is only one possible state for the outcome. We can thus draw the simple identity shown in Fig. 7.19

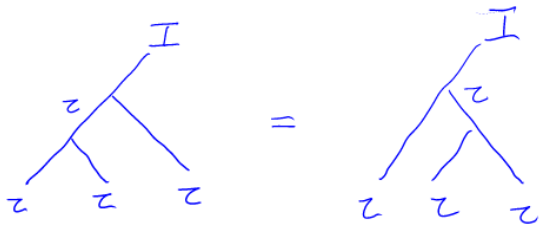
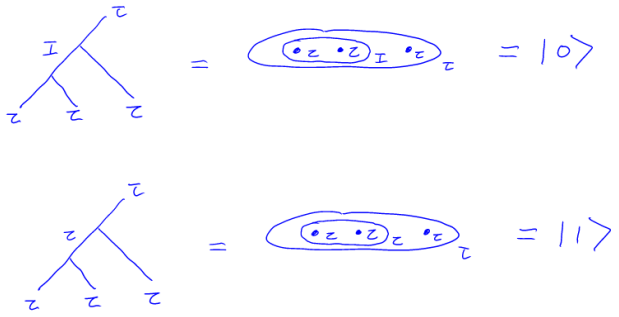
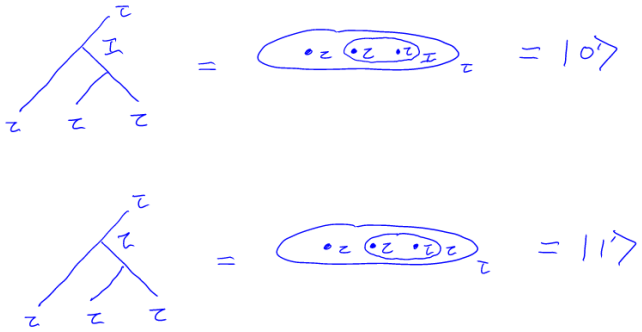


Figure 7.19: There is only one state in the Hilbert space of three fibonacci anyons fusing to the identity. Thus it does not matter if you fuse the left two first or the right two first, you are describing the same state.

The more interesting situation is the case where the three Fibonacci anyons fuse to τ. In this case, there is a two dimensional space of states, and this two dimensional space can be described in two ways. We can fuse the left two particles first to get either I (yielding overall state $|0\rangle$) or to get τ (yielding overall state $|1\rangle$). See the top of Fig. 7.20. On the other hand, we could fuse the right two particles first to get either I (yielding overall state $|0'\rangle$) or to get τ (yielding overall state $|1'\rangle$). See the bottom of Fig. 7.20.



Fusing The Two Particles on the Left First



Fusing The Two Particles on the Right First

Figure 7.20: Two ways to describe the same two dimensional space. The basis $\{|0\rangle, |1\rangle\}$ fuses the left two particles first, whereas the basis $\{|0'\rangle, |1'\rangle\}$ fuses the right two particles first.

The space of states spanned by the three anyons is the same in either description. Thus, there

must be a unitary basis transform given by

$$\begin{pmatrix} |0\rangle \\ |1\rangle \end{pmatrix} = \begin{pmatrix} F_{00'} & F_{01'} \\ F_{10'} & F_{11'} \end{pmatrix} \begin{pmatrix} |0'\rangle \\ |1'\rangle \end{pmatrix} \tag{7.2}$$

Thus here F is a two by two matrix, and in the notaion of the F matrix defined in Fig. 7.18, this two by two matrix is $[F^{\tau}_{\tau\tau\tau}]_{ab}$ and the indices a, b should take the values I and τ instead of 0 and 1, but it is perhaps easier to use the notation shown here for more clarity.

For the Fibonacci theory the F matrix is given explicitly by

$$F = \begin{pmatrix} \phi^{-1} & \phi^{-1/2} \\ \phi^{-1/2} & -\phi^{-1} \end{pmatrix}$$

where $\phi^{-1} = (\sqrt{5}-1)/2$. As one should expect for a change of basis, this matrix is unitary. Below in the next section we will discuss how this matrix is derived.

It is important to emphasize that the F-matrix is the same even if one of the anyons charges being fused is actually a cluster of several anyons. For example, in Fig. 7.21, this is precisely the same transformation as in Eq. 7.2, but we must view the cluster of two anyons on the left (underlined in red), which fuse to τ as being a single τ particle.

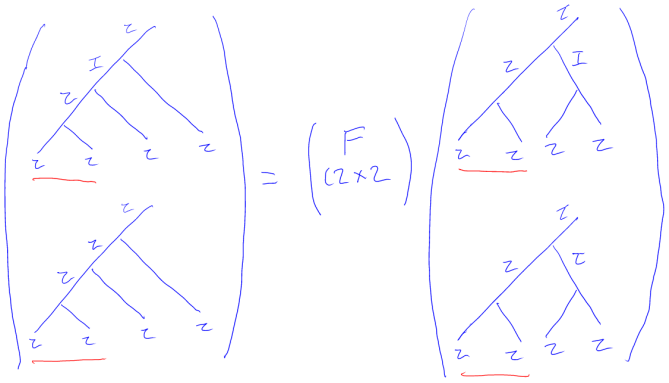


Figure 7.21: The F-matrices are the same even if one of the anyon charges is made up of a cluster of other anyons. In this particular picture, the cluster of two anyons on the left (underlined in red) has charge τ. if one were to replace this with just a single τ, this would be precisely the same transformation as in Eq. 7.2.

7.5.1 Pentagon

It is possible to describe the same Hilbert space in many ways. For example, with three anyons, as in Fig. 7.13, one can describe the state in terms of the fusion channel of the two anyons on the left, or in terms of the two on the right. Ie., we can describe $(a \times b) \times c$ or $a \times (b \times c)$ And as in Fig. 7.18, these two descriptions can be related via an F-matrix.

When there are four anyons, there are still more options of how we group particles to describe the states of the Hilbert space, and these can also be related to each other via F matrices similarly, as shown in Fig. 7.21. The fact that we can should be able change the connectivity of these tree diagrams then allows one to make multiple changes in the trees as shown in Fig. 7.22 (the step in the furthest upper left is equilvalent to that shown in Fig. 7.21). Indeed, in this figure one sees that one can go from the far left to the far right of the diagram via two completely different paths (the top and the bottom path) and the end result on the far right should be the same either way.

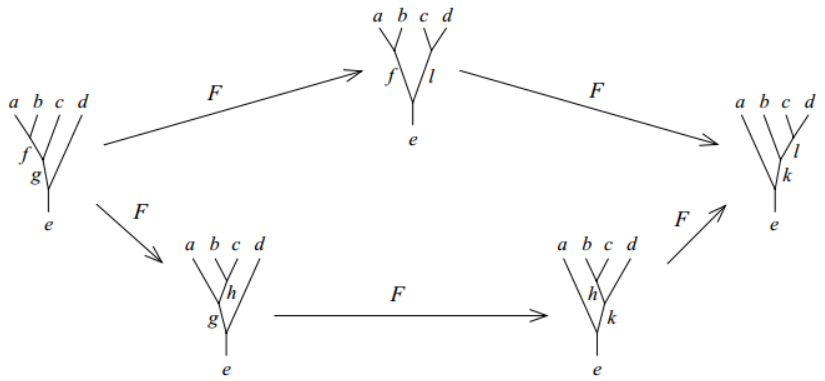


Figure 7.22: Pentagon Diagram. Figure Stolen from Bonderson's thesis.

This diagram, known as the pentagon diagram, puts a very strong contraint on the F-matrices, which written out algebraically would be

$$[F_e^{fcd}]_{gl}[F_e^{abl}]_{fk} = \sum_h [F_g^{abc}]_{fh}[F_e^{ahd}]_{gk}[F_k^{bcd}]_{hl}$$

where the left hand side represents the top route of the figure and he right hand side represents the bottom route.

For very simple theories, such as the Fibonacci anyon theory, the Pentagon diagram is sufficient to completely define the F-matrices (up to some gauge convention choices). This is a nice exercise to try!

One might think that one could write down more complicated trees and more complicated paths through the trees and somehow derive additional constraints on the F-matrices. A theorem by MacLane, guarantees that no more complicated trees generate new identities beyond the pentagon diagram.

Chapter 8

Braiding and Twisting

8.1 Twists

In an Anyon theory (or topological quantum field theory in general) each particle a is endowed with a **topological spin**, or **conformal scaling dimension**, usually called h_a related to the **twist factor** θ_a

$$\theta_a = e^{2\pi i h_a}$$

We note that in many cases quantities of interest will depend only on the twist factor, i.e., the fractional part of the topological spin $h_a \mod 1$. It is often hard to pin down the value of the topological spin itself.

In our diagrammatic notation, we have

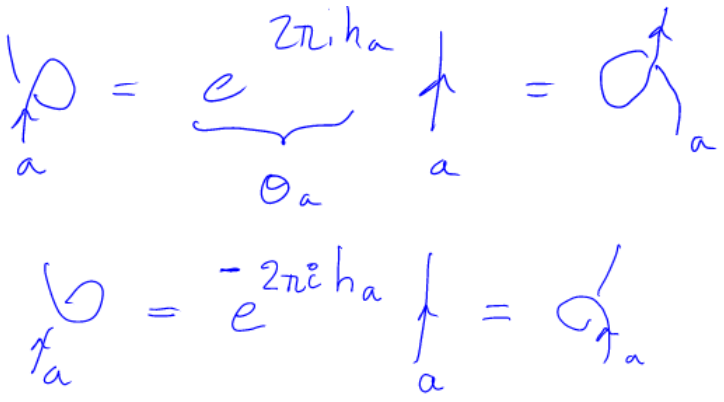


Figure 8.1: Twist Factor

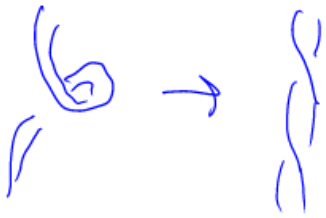


Figure 8.2: Pulling Straight

Recall, we should treat particle world-lines as ribbons, so that a little loop of this type can be pulled straight as in Fig. 8.2 to represent a particle twisting around its own axis, as well as giving the phase of exchange for two identical particles. Two cases are well known to us if the spin h_a is an integer, then $e^{2\pi i h_a}$ is the identity, and this particle is a boson. If h_a is a half-odd-integer, then the phase is -1 and the particle is a fermion. The vacuum, or identity particle, should have zero scaling dimension, $h_I = 0$.

8.2 R-matrix

Consider the possibility of two particles fusing to a third as shown in Fig. 8.3

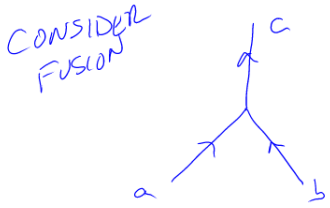


Figure 8.3: Two particles fusing to a third. For this anyon system $a \times b = c + \ldots$, and c is the particular fusion channel that has occured in this diagram.

We have $a \times b = c + \ldots$. I.e., c is among the possibile fusion channels that can occur and we assumes in the diagram that c is the particular fusion channel that has occured. Now let us consider braiding a and b around each other before fusing them as in Fig. 8.4. This diagram defines the so-called R-matrix. Here we have dropped the arrows and we show the particle world lines as ribbons to show that there are no additional self-twists.

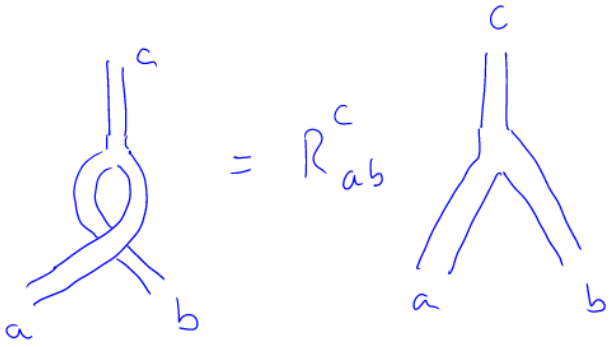


Figure 8.4: Definition of R-matrix. Here we drop the arrows for convenience of notation and we draw the particle world lines as ribbons to show that no additional self-twists are incurred by the particles.

Note that braiding anything with the identity particle should be trivial.

To see the relationship between braiding and twisting, consider applying the R matrix twice to make a double twist as in Fig. 8.5. By pulling tight the double twist, the diagram can be reduced to twist factors previously defined, and this fixes R^c_{ab} up to a possible minus sign.

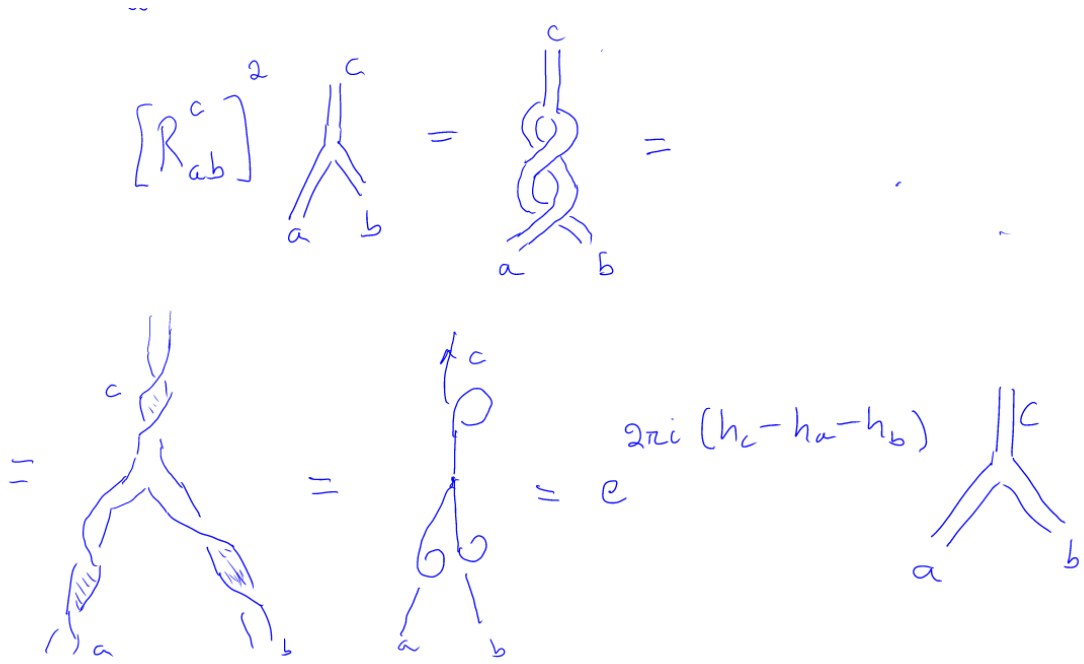


Figure 8.5: Relation of R-matrix to twist factors.

Which we can generally write as

$$[R^c_{ab}]^2 = e^{2\pi i(h_c - h_a - h_b)} = \theta_c/(\theta_a\theta_b)$$

Example: Fibonacci Anyons. In the Fibonacci theory, two τ particles can fuse to either τ or

I. Applying the above relationship, we have

$$[R^{\tau}_{\tau\tau}]^2 = e^{2\pi i(h_\tau - h_\tau - h_\tau)} = e^{-2\pi i h_\tau} \tag{8.1}$$

$$[R^{I}_{\tau\tau}]^2 = e^{2\pi i(h_I - h_\tau - h_\tau)} = e^{-4\pi i h_\tau} \tag{8.2}$$

Using the F and R matrices for a general anyon theory we can evaluate the unitary transform associated with any braid. Recall the two possible states of three τ particles fusing to τ as shown in Fig. 8.6.

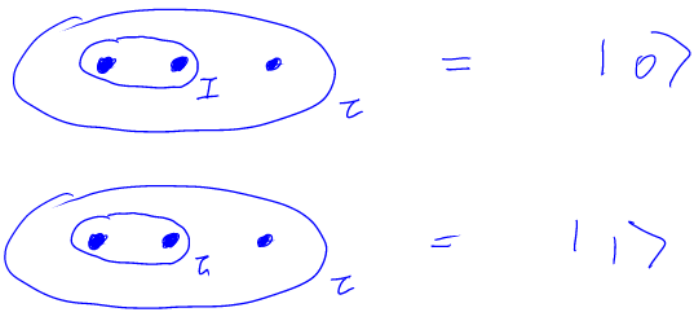


Figure 8.6: The two stats of three τ particles fusing to τ. Unmarked dots are τ particles.

Now consider braiding the two leftmost particles around each other.

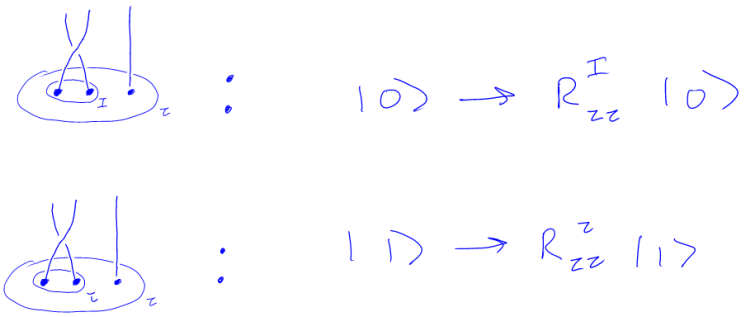


Figure 8.7: Braiding the two left particles in this basis gives a phase dependenet on the fusion channel of the two particles.

The result of this fusion gives a phase, either $R^I_{\tau\tau}$ if the fusion channel of the two particles is I or $R^\tau_{\tau tau}$ if the fusion channel of the two particles is τ.

Note that the braiding operator is a linear quantum mechanical operator, so it acts on superpositions.

$$R : (\alpha|0\rangle + \beta|1\rangle) = \alpha R^I_{\tau\tau}|0\rangle + \beta R^\tau_{\tau\tau}|1\rangle$$

This is what is known as a controlled phase gate in quantum information processing – the phase accumulated depends on the state of the qubit.

Now how can we evaluate the braid shown in Fig. 8.8? The trick here is to use the F-matrix to change the basis such that we know the fusion channel of the right two particles, and then once we know the fusion channel we can use the R-matrix, and if we want we can then use the F-matrix to transform back to the original basis.

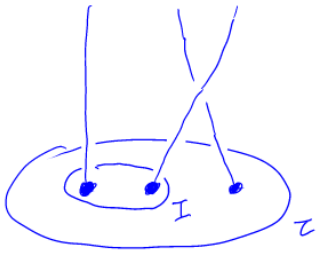


Figure 8.8: How does one evaluate this braid?

To see how this works, Recall that we can use the F matrix to write (See Eq. 7.2)

$$|0\rangle = F_{00'}|0'\rangle + F_{01'}|1'\rangle$$

or in diagrams (see Fig. 8.9).

Figure 8.9: The F-matrix relation in diagram form. See Eq. 7.2

On the right hand side (i.e., in the prime basis) we know the fusion channel of the rightmost two particles, so we can braid them around each other and use the R-matrix to accumulate the corresponding phase.

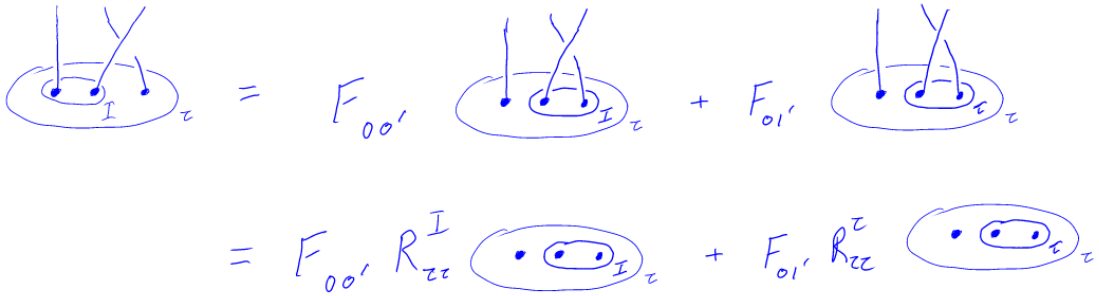


Figure 8.10: To braid particles, switch basis until we know the fusion channel of the two particles we want to braid, and then we can apply the F-matrix.

To describe this in equations, we can write the operator that braids the rightmost two particles

as R_{23} and then we have

$$\begin{aligned} R_{23}|0\rangle &= R_{23}(F_{00'}|0'\rangle + F_{01'}|1'\rangle) \\ &= F_{00'}R_{23}|0'\rangle + F_{01'}R_{23}|1'\rangle \\ &= F_{00'}R^I_{\tau\tau}|0'\rangle + F_{01'}R^\tau_{\tau\tau}|1'\rangle \qquad (8.3) \\ &= F_{00'}R^I_{\tau\tau}\left([F^{-1}]_{0'0}|0\rangle + [F^{-1}]_{0'1}|1\rangle\right) \qquad (8.4) \\ &\quad + F_{01'}R^\tau_{\tau\tau}\left([F^{-1}]_{1'0}|0\rangle + [F^{-1}]_{1'1}|1\rangle\right) \\ &= \left(F_{00'}R^I_{\tau\tau}[F^{-1}]_{0'0} + F_{01'}R^\tau_{\tau\tau}[F^{-1}]_{1'0}\right)|0\rangle \\ &\quad + \left(F_{00'}R^I_{\tau\tau}[F^{-1}]_{0'1} + F_{01'}R^\tau_{\tau\tau}[F^{-1}]_{1'1}\right)|1\rangle \end{aligned}$$

Where between Eq. 8.3 and 8.4 we have used the inverse F transform to put the result back in the original $|0\rangle$ and $|1\rangle$ basis. Note that for this particular case F and F^{-1} are the same matrix (however we write out the inverse explicitly for clarity!).

This general principle allows us to evaluate any braiding of particles. We always convert to a basis where the fusion channel of the two particles to braided is known, then we apply the R matrix directly. At the end we can transform back to the original basis if we so desire.

8.3 The Hexagon

As with the case of the F-matrix, there are strong consistency constraints on the R-matrices given a set of F-matrices (indeed, it is possible that for a given set of F-matrices that satisfy the pentagon, there may not even exist a set of consistent R-matrices!). The consistency equations are known as the hexagon equations and are shown diagrammatically in Fig. 8.11

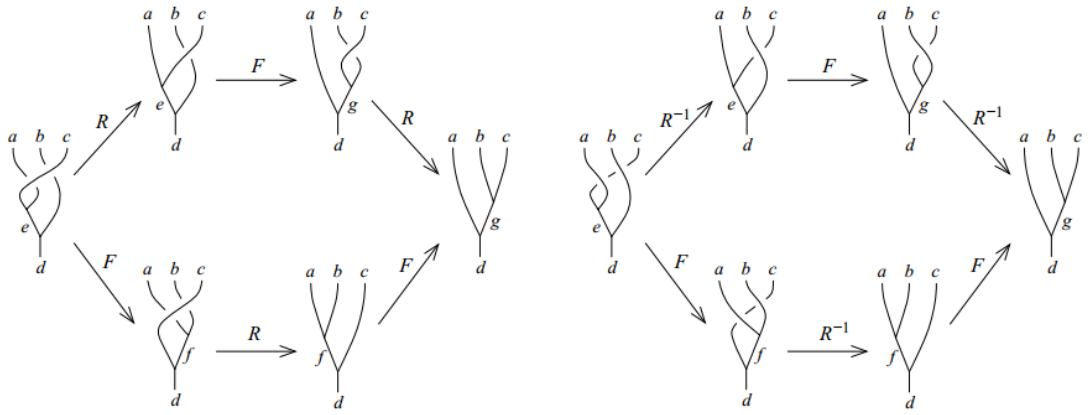


Figure 8.11: The hexagon equations in graphical form. (nice picture stolen from Bonderson thesis)

In equations this can be expressed as

$$R^{ca}_e[F^{acb}_d]_{eg}R^{cb}_g = \sum_f [F^{cab}_d]_{ef}R^{cf}_d[F^{abc}_d]_{fg}$$

$$[R^{ca}_e]^{-1}[F^{acb}_d]_{eg}[R^{cb}_g]^{-1} = \sum_f [F^{cab}_d]_{ef}[R^{cf}_d]^{-1}[F^{abc}_d]_{fg}$$

The top equation is the left diagram whereas the lower equation is the right diagram in Fig. 8.11. The left hand side of the equation corresponds to the upper path, whereas the right hand side of the equation corresponds to the lower path.

In simple theories such as the Fibonacci theory (knowing the F matrices) the Hexagon equation, almost uniquely defines the R-matrices. In fact there are two consistent solutions to the Hexagon equations for the Fibonacci theory.

$$R^{\tau}_{\tau\tau} = e^{\pm 3\pi i/5}$$
$$R^{I}_{\tau\tau} = e^{\mp 4\pi i/5}$$

These two solutions correspond to left and right handed versions of the Fibonacci theory corresponding to twist factors for the elementary Fibonacci anyon of.

$$\theta_\tau = e^{\pm 4\pi i/5}$$

8.4 Ocneanu Rigidity

Given a set of fusion rules, the pentagon and hexagon equation are very very strong constraints on the possible F and R matrices that can result. (For example, as mentioned above, with Fibonacci fusion rules, there is only one solution of pentagon and then only two solutions the hexagon). In fact, it is a general principle that the pentagon and hexagon for any set of fusion rules for a finite set of particles will have a finite set of solutions. In particular, once we have a set of solutions, in no sense is there a way that we can deform the values of F and R by a small amount and have another solution. This is a principle known as *rigidity* of the solutions, and it was first pointed out by Ocneanu. This principle makes it possible to contemplate putting together a sort of "periodic table" of possible anyon theories, starting with those having very few particle types. In fact, such periodic tables have been compiled up to about 5 or 6 different particle types (See Rowell Strong and Wang[1]). There is nothing in principle that prevents one from listing all the possible anyon theories even for more particle types although the search for all solutions becomes extremely difficult for greater numbers of particles.

[1]It is often useful to impose one more condition, that the theory is "modular" which we will discuss below in section ***. Most well behaved theories are modular, although the presence of a fermion makes a theory non-modular — indicating how difficult it is to properly treat fermions! As far as we can tell from the known periodic table, all modular theories can be described in terms of some sort of Chern-Simons theory or closely related construction!

Chapter 9

Kauffman Categories

Good reference is the book by Kauffman,

To give a definite example of an anyon theory, let us back up to the Kauffman invariant. What was missing in that picture was the idea of multiple particle types and fusion. Here we try to construct fusion rules based on the Kauffman rules of Fig. 2.2. In fact, to begin with we don't even want to consider braiding, just fusion. So for now we can neglect the braiding rule and focus only on the loop rule shown in Fig. 9.1.

Figure 9.1: The loop rule for the Kauffman invariant and the Temperly-Lieb algebra.

If we consider an algebra of loops (using the loop rule that one loop gets a value d) — but no braiding allowed — this algebra is known as the Temperly-Lieb algebra.

Now we would like to ask whether we can fuse two of these strings together to make another particle. Since we are going to construct multiple particle types, let us call the basic string the particle type "1" and the vacuum will be denoted by the particle type "0".

One possibility is to fuse the two particles to the vacuum as shown in Fig. 9.2

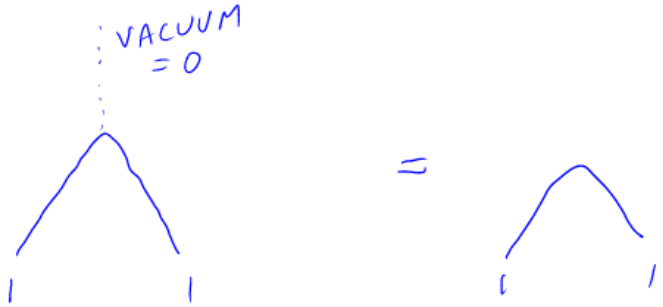


Figure 9.2: Fusing two 1-particles to the vacuum

The fact that two 1-particles can fuse to the vacuum tells us immediately that 1 is its own antiparticle.

$$1 = \bar{1}$$

We might also consider the possibility that two of these 1-particles can fuse to something besides the vacuum, in a way similar to that shown in Fig. 9.3.

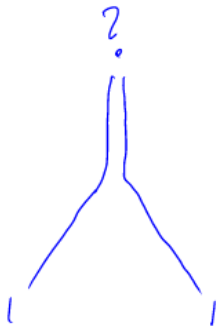


Figure 9.3: Attempting to Fuse two 1-particles to something different from the vacuum

This is a good idea, but it isn't yet quite right. We need to assure that if we have two different particle types they are appropriately "orthogonal" to each other. This orthogonality must be in the sense of the locality, or no-transmutation rule. A particle type must not be able to spontaneously turn into another particle type (without fusing with some other particle or splitting). To help us construct particles which will obey this rule, we must construct projection operators.

9.1 Jones-Wenzl Projectors

The definition of a projector is an operator P such that $P^2 = P$, so it has eigenvalues 0 and 1. We would like to construct projectors out of strings.

To begin with, we will construct projectors out of two incoming 1-particles (two elementary strings). We can construct the projector P_0 that forces the two incoming particles to fuse to the vacuum.

Figure 9.4: The projector of two strings to the vacuum P_0.

We now need to check that $P_0^2 = P_0$. To apply the P_0 operator twice we connect the strings coming out the top of the first operator to two strings coming in the bottom of the second operator. As shown in Fig. 9.5, using the fact that a loop gets value d we see that $P_0^2 = P_0$ meaning that P_0 is indeed a projector.

Figure 9.5: Checking that $P_0^2 = P_0$.

We now consider the possibility that two strings near each other can fuse to something else, which we call P_2. Let us define $P_2 = I - P_0$ where I is the identity operator, i.e, just two parallel strings. Diagrammatically we have Fig. 9.6

Figure 9.6: The projector of two strings to the nontrivial particle $P_2 = I - P_0$.

We can then algebraically check that this is indeed a projector

$$P_2^2 = (I - P_0)(I - P_0) = I - 2P_0 + P_0^2 = I - P_0 = P_2$$

and also we can check that it is orthogonal to P_0, by

$$P_0 P_2 = P_0(I - P_0) = P_0 - P_0^2 = 0$$

and similarly to show $P_2 P_0 = 0$.

Often it is convenient to draw these projection operators as a labeled box, as shown in Fig. 9.7.

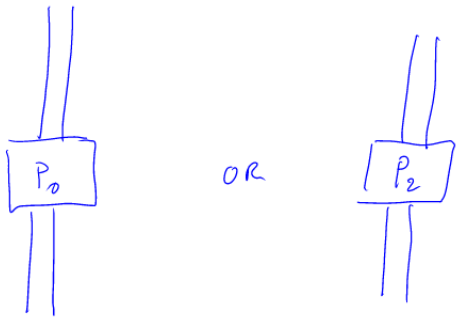


Figure 9.7: Drawing the two possible fusion channels of two strings as a box labeled P_0 or P_2

Sometimes we simply draw a single line with a label, 0 or 2 respectively.

Abelian Case: In the case where $d = \pm 1$ we proved for Homework that two horizontal lines equals $\pm$ two vertical lines as shown in Fig. 9.8. In this case, notice that the projector $P_2 = 0$ (the two terms in the projector are equal with opposite signs).

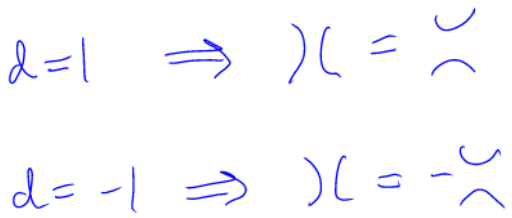


Figure 9.8: Two cases where the Kauffman invariant rules become very simple. If you have not convinced yourself of these rules, try to do so!. Note that $d = 1$ occurs for bosons or fermions and $d = -1$ occurs for semions.

The theories in queston here are bosons or fermions for $d = 1$ and semions for $d = -1$. All of these theories are abelian, so it is not surprising that two particles that can fuse to the identity cannot fuse to another species as well. We have only the fusion to the vacuum shown in Fig. 9.2. Thus the entire fusion rules of these theories are

$$1 \times 1 = 0$$

where again 0 is the identity or vacuum.

General Case:

For other values of d, however, two strands (each labeled 1) can fuse either to 0 or to 2 as shown in Fig. 9.9.

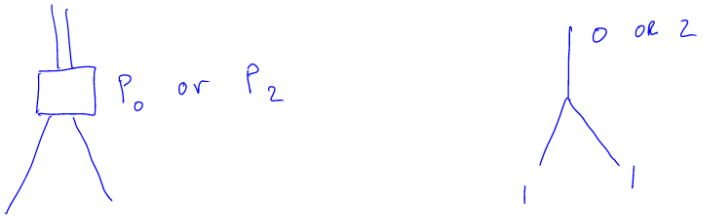


Figure 9.9: Possible fusions of two strands

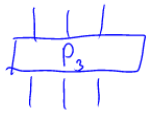
We can write the fusion rule

$$1 \times 1 = 0 + 2$$

We might ask whether it is possible to assemble a third type of particle with two strands. It is obvious this is not possible since $P_0 + P_2 = I$ forms a complete set.

Three Strands:

We can move on and ask what kind of particles we can make if we are allowed to fuse three strands together. We want to try to construct a three leg projector of the form in Fig. 9.10.



General Form:

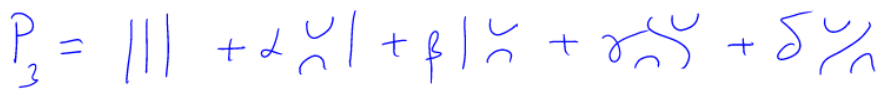
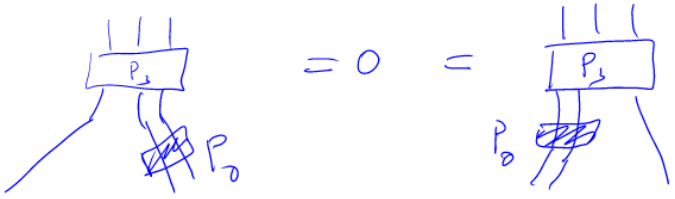


Figure 9.10: A three-leg Jones-Wenzl projector. The most general possible form of P_3 is written in the lower line.

We should certainly enforce that $P_3^2 = P_3$ so that this acts as a projector. However, there are other things we want to enforce as well. We want $0 \times 1 = 1$ which means we should not be able to fuse P_0 with a single strand to get P_3. Diagrammatically we insist on relations like Fig. 9.11.



or Equivalently

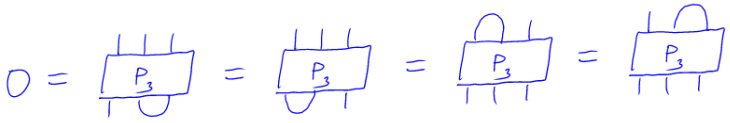


Figure 9.11: Insisting that 0×1 does not give 3

However, we should alllow fusions of the form $1 \times 2 = 3$ as shown in Fig. 9.12

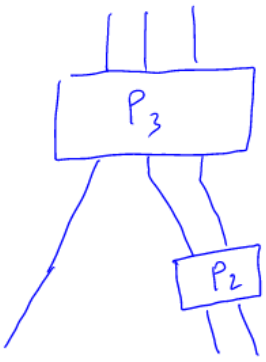


Figure 9.12: We allow $1 \times 2 = 3$

Enforcing the condition in Fig. 9.11, along with $P_3^2 = P_3$ gives the form of P_3 shown in Fig. 9.13.

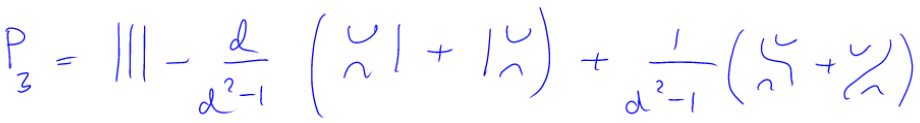


Figure 9.13: Form of the P_3 projector in terms of the parameter d.

(deriving this is a good exercise!).

Ising Anyons

Consider the case where $d = \sqrt{2}$. Here it is possible to show that P_3 vanishes when evaluated in any diagram (this is also a good exercise, try it!). It is similarly possible to show that $P_4 = 0$ and so forth. Thus, in this theory there are only three particle types P_0, P_1 and P_2. We have $2 \times 2 = 0$ as shown in Fig.9.14 and $2 \times 1 = 1$ as shown in Fig. 9.15. (Note that showing $2 \notin 2 \times 2$ requires another explicit calculation, not shown here!)

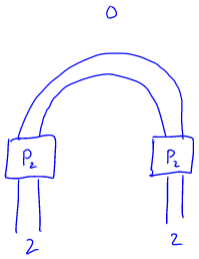


Figure 9.14: $2 \times 2 = 0$.

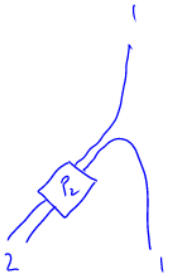


Figure 9.15: $2 \times 1 = 1$. We recognize this as the fusion $1 \times 1 = 2$ from Fig. 9.9 just turned on its sid.e

We then have the full set of nontrivial fusion rules

$$1 \times 1 = 0 + 2$$
$$2 \times 2 = 0$$
$$1 \times 2 = 1$$

which we recognize as Ising fusion rules (see 7.2.2) where $1 = \sigma$ and $2 = \psi$.

Note: It is not coincidence that the quantum dimension of the nonabelian particle (the 1-particle, or σ) is $\sqrt{2}$ (see section '7.2.2), and that $d = \sqrt{2}$ as well. It is a general principle that the value of a loop will be the same as the quantum dimension of the corresponding particle, as we will see below.

9.2 F-matrices

We can determine the F-matrices directly from the graphical algebra. Let us consider the case of 3 single strands coming in the bottom and fusing to a single strand going out the top. I.e, we are looking at the matrix F^1_{111}. The F matrix is nontrivial since there is more than one fusion channel when we fuse the 1's together: $1 \times 1 = 0 + 2$. Let us write the F-matrix as

$$F = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$

by which we mean the diagram Fig. 9.16

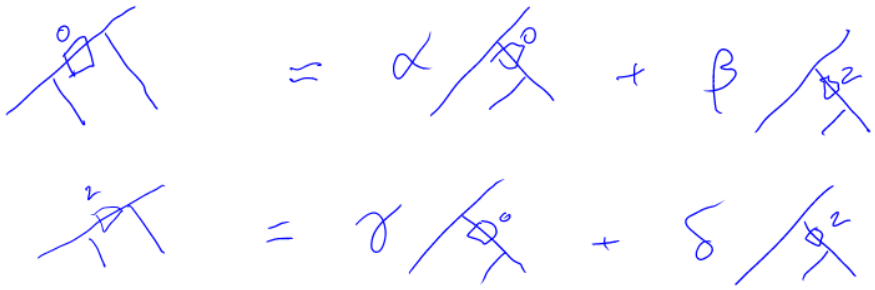


Figure 9.16: An F-move. The boxes are either P_0 or P_2 projectors.

We now realize that graphically a P_0 projector is $1/d$ times a simple turnaround, whereas a P_2 projector is the sum of two terms, the first being just two strands going parallel (i.e,, the identity) and the second being $-P_0$. (See Figs. 9.4 and 9.6). We can then write graphically Fig. 9.17

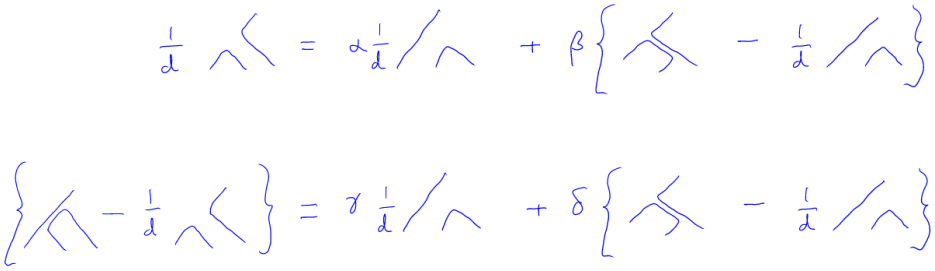


Figure 9.17: Plugging in the form of the projectors.

We then match up terms. In the first line we see the the diagram on the left is topologically like the first term in the brackets on the right, so we have $\beta = 1/d$. Similarly the first term on the right is the same as the second term in the brackets, so $\alpha = \beta = 1/d$. Then in the second line we have the second term in brakets on the left is the same as the first term in brackets on the right, so we have $\delta = -1/d$. Then the remaining terms, the first term in brakets on the left, the first term on the right, and the secon term in brackets on the right, are all the same shape, so we have $1 = \gamma/d - \delta/d$ or $\gamma = d - 1/d$. Thus obtaining the form of the F matrix

$$F = \begin{pmatrix} 1/d & 1/d \\ d - 1/d & -1/d \end{pmatrix}$$

9.3 Twisting and Braiding

So far we have not yet used the braiding rules of the Kauffman invariant, we have only used the loop rule. We finally can reintroduce the braiding rules for the Kauffman invariant for evaluating crossings as in Fig. 2.2. As shown in Fig. 2.5, comparing to Fig. 8.1 we see that the twist factor of the single strand is $\theta_1 = -A^{-3}$. It is a reasonably straightforward exercise to use these crossing rules to evaluate the twist factors for other particles in the theory, as well as the R-matrices. Just to do a simple example, let us evaluate R^2_{11} shown in Fig. 9.18

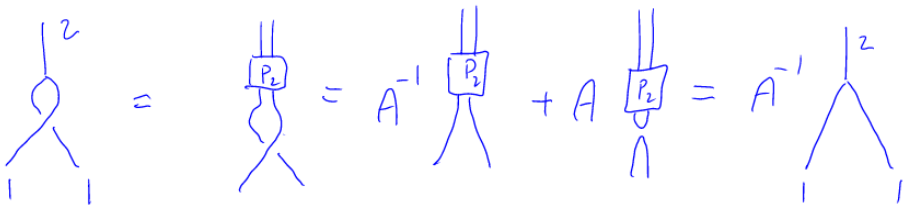


Figure 9.18: Evaluation of $R^2_{11} = A^{-1}$.

Here the term with the coefficient A vanishes because of the orthogonality of P_2 and P_0.

Chapter 10

Diagrammatic Algebra, the S-matrix and the Verlinde Relation

We have built up our anyon theories and now, using F and R matrices we can generally figure out how the degenerate Hilbert space $V(\Sigma)$ evolves (where by Σ we mean a surface with particle in it).

We are almost at the point where we have a full diagramatic calculus — which would produce a number as an output given any world-lines as input

$$Z(\text{Manifold with particle world lines in it}) \to \mathbb{C}$$

Note that while diagrammatic calculus for the Kauffman case is often quite simple, there can be some nasty bookkeeping glitches for other anyon theories. For careful details of how all of the details, see Kitaev 2005 or Bonderson thesis.

First, we should be careful about our normalization when we evaluate some knot or link of world lines[1]. We choose our evaluation of a world line link to be of the form

$$\langle \text{Link} \rangle \equiv \frac{Z(S^3 \text{ with embedded Link})}{Z(S^3)} = Z(S^2 \times S^1 \text{ with embedded Link})$$

where in the case of $S^2 \times S^1$ we require that the Link not go around the nontrivial handle of the S^1. This normalization is chosen so that the evaluation of the empty link will give unity (as discussed in chapter 6).

By using F's and R's we hope to reduce diagrams to a collection of non-linking labeled loops (labeled with their particle type), similar to what we did in evaluating the Kauffman invariant. We then need to know what value to give a particular loop.

10.1 Normalizations and Loops

Let us define $d_a > 0$ to be the value associate with the a loop of particle of type a as shown in Fig. 10.1[2] These quantities will turn out to be the quantum dimensions of the particles, but we have not shown this yet!

[1] We allow branching world lines which correspond to fusion or splitting

[2] In some cases it is convenient to define the value of a loop to be negative, as in the case of the semion Kauffman theory discussed above. However, by redefining some F-matrix elements, one can always work with the convention that d_a is positive, although this comes at the expense of having troublesome minus signs pop up in other places! These minus signs are known as Frobenius-Schur indicators.

Figure 10.1: A loop of particle type a is given value $d_a > 0$. This will turn out to be the quantum dimension of the particle.

We have not yet decided what value this loop should get. However, we can look back to 6.5 to note that we have

$$Z(S^3; a \text{ loop linking } b \text{ loop}) = S_{ab}$$

where S_{ab} is the unitary matrix known as the modular S-matrix. Recall that S should be unitary because it can be interpretated as a change of basis. (Theories where the S matrix comes out non-unitary are considered badly behaved, or "non-modular". We will ignore this harder case for now!).

We can then think of the single loop d_a as particle a linking the vacuum, so we write

$$Z(S^3; a \text{ loop}) = S_{a0} = S_{0a}$$

and further we can write the normalizing factor $Z(S^3)$ as vacuum linking vacuum, so we have the value of a single loop as

$$d_a = S_{a0}/S_{00}$$

The fact that S is unitary gives us a useful identity

$$1 = \sum_a |S_{a0}|^2 = |S_{00}|^2 \sum d_a^2$$

which gives us

$$Z(S^3) = S_{00} = 1/\mathcal{D}$$

where $\mathcal{D}$ is known as the total quantum dimension and is given by

$$\mathcal{D}^2 = \sum_a d_a^2$$

Note that, as of this point we still have not shown that the d_a's, i.e., the values of the loops, are related to the quantum dimensions.

10.2 Quantum Dimensions

Now, we claim that these loop quantites d_a should satisfy the fusion algebra

$$d_a d_b = \sum_c N_{ab}^c \, d_c \tag{10.1}$$

or diagrammatically we have Fig. 10.2

Figure 10.2: The quantum dimensions satisfy the fusion algebra.

This rule seems rather natural, that a and b can fuse together to form c in all possible ways. However, to prove it is a bit more complicated than this argument, and is given in the appendix to this chapter.

Now, given Eq. 10.1, if we think of the fusion multiplicity for particle a, as a matrix N_a with indices b and c, and we think of d_c as a vector $\vec{d}$ we can write

$$d_a \vec{d} = [N_a]\vec{d}$$

I.e, the vector $\vec{d}$ is an eigenvector of N_a with eigenvalue d_a.

Note that the matrix N_a has only non-negative elements and $\vec{d}$ has only positive elements. This allows us to apply the Perron-Frobenius theorem which says that for matrices with only non-negative elements[3] there is a unique eigenvector with all positive entries, and it corresponds to the largest eigenvalue. Thus we conclude that d_a is actually the largest eigenvalue of the matrix N_a and it has eigenvector $\vec{d}$.

Recall that our previous definition of the quantum dimension d_a is that it is the largest eigenvalue of the fusion multiplicity matrix N_a. Thus we have rigorously shown that the value d_a of the loop in the graphical algebra is precisely the quantum dimension!

10.3 Verlinde Algebra

Using the locality principle (or no-transmutation) principle (See Fig. 7.7) we can show that a closed loop of type a around a world line of type x gives some constant which we call $\tilde{S}_{ax}$ as shown in Fig. 10.3.

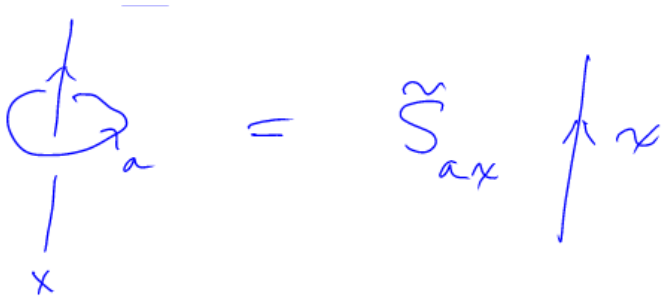


Figure 10.3: The locality priniciple tells us that the value of a loop around a world line is some number which we call $\tilde{S}_{ax}$

[3]Actually the simplest version of Perron-Frobenius requires all positive elements. Using the theorem for non-negative matrices allows there to be a second eigenvalue with same magnitude but opposite sign — this does not change the conclusion.

by bending the top of x and forming a closed loop with the bottom of x, we construct linked rings on the left of this equation which we relate to the modular S-matrix, but on the right we form just a single x-loop.

$$S_{ax} = Z(S^3,\ a \text{ loop links } x \text{ loop}) = \tilde{S}_{ax} Z(S^3,\ x\text{-loop}) = \tilde{S}_{ax} S_{0x}$$

from which we conclude

$$\tilde{S}_{ax} = \frac{S_{ax}}{S_{0x}} \tag{10.2}$$

On the other hand, if we have two loops a and b around x, we can fuse the two loops to all possible loops c as shown in Fig.10.4. This identity is entirely analogous to that of Fig. 10.2, and the rigorous derivation is given in the appendix.

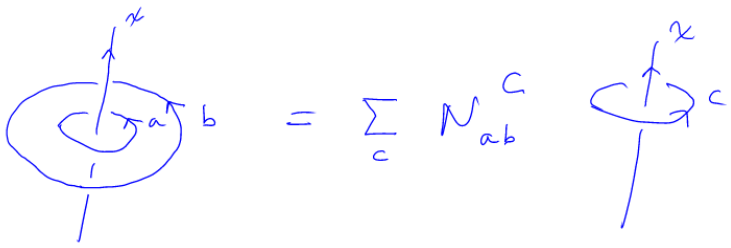


Figure 10.4:

On the other hand, we could also evaluate the left hand side of Fig. 10.4 by applying the identity of Fig. 10.3 twice in a row, and similarly we can evaluate the right hand side of Fig. 10.4 by applying Fig. 10.3once. Thus we obtain the identity

$$\tilde{S}_{ax}\tilde{S}_{bx} = \sum_c N_{ab}^c \tilde{S}_{cx}$$

This important result can be re-presented in two important ways. First, inverting this matrix equation gives

$$N_{ab}^c = \sum_x \tilde{S}_{ax}\tilde{S}_{bx}[\tilde{S}^{-1}]_{xc}$$

Plugging in the value of $\tilde{S}$ from Eq. 10.2, and using the fact that the modular S matrix is unitary, we obtain the famous Verlinde formula

$$N_{ab}^c = \sum_x \frac{S_{ax}S_{bx}S_{xc}^*}{S_{0x}}$$

which tells us that all the information about the fusion algebra is contained entirely within the modular S matrix!

A second way to present this important results is to write it in the form

$$[S^\dagger N_a S]_{xy} = \tilde{S}_{ay}\delta_{xy}$$

where N_a here is the matrix N_{ab}^c with indices b and c. Thus the result tells us that the modular S matrix is precisely the unitary diagonalizing matrix we were looking for in Eq. 7.1!

10.4 Return of Kirby Color

As mentioned in section 6.4.2, one can assemble a string called the "Kirby Color" (or Ω string) that is the sum of all strings weighted by the S-matrix.

$$|\Omega\rangle = \sum_a S_{0a}|a\rangle = \frac{1}{\mathcal{D}}\sum_a d_a|a\rangle$$

This string has some remarkable properties. Suppose we loop this string around a string x similar to that of Fig. 10.3. The result then looks like

$$\sum_a S_{0a}\tilde{S}_{ax}|x\rangle = \mathcal{D}\delta_{x0}|0\rangle$$

where we have used the fact that S is unitary, that $S_{0a} = S_{a0}$ is real, and that $S_{00} = 1/\mathcal{D}$. This is shown explicityly in fig. 10.5

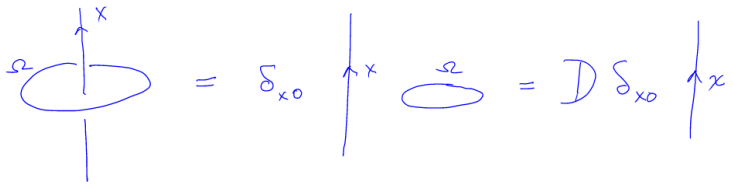


Figure 10.5: The kirby color string projects to the vacuum going through it

Thus, a loop of Kirby color string projects to zero (or vacuum) flux going through it! This principle is extremely useful in later attempts to construct topological models.

Further, the Kirby color string can be used, as mentioned in section 6.4.2 to build up a manifold invariant from anyon braiding rules. Indeed we can check this. The evaluation of the empty knot is defined to be $1 = Z(S^3)/Z(S^3)$. Surgery on a single loop takes S^3 to $S^2 \times S^1$. And evaluation of a single loop of Kirby color gives $\mathcal{D} = Z(S^2 \times S^1)/Z(S^3)$. So this appears to be working! One should be a bit careful with this because one needs to properly account for twists in loops which we have not done here.

10.5 S and θ and that is all?

In building up an anyon theory, we now have compiled a larege amount of data. Say there are M particle types, then we have F matrices, which have 6 indices, each running from 1 to M, we have N matrices with three indices, we have R matrices with three indices, we have S matrices with two indices, and d's and θ's. This seems like a huge amount of data needed to keep track of (and in some sense it is a huge amount of data). However, due to the idea of Rigidity, it is believed that you need only specify the matrix S_{ab} and the values of the twists θ_a and you completely pin down the rest of the theory. I do not believe this statement is proven, but there are no counter-examples known.

10.6 Appendix: Quantum Dimensions Satisfy the Fusion Algebra

We would like to show the identity shown in Fig. 10.2. We need a few useful pieces. First note that we can use an F-move on parallel lines to show the identity shown in Fig. 10.6.

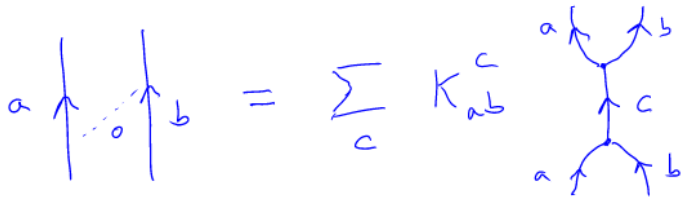


Figure 10.6: An F-move. If a and b do not fuse to c, then the coefficient κ_{ab}^{c} must be zero. And if a and b do fuse to c then κ_{ab}^{c} is not zero. Note that the constant κ_{ab}^{c} shown here is typically notated as F_{abc}^{ab0}. This is quite similar to $[F_{ab\bar{b}}^{a}]_{0c}$ except that some lines pointing up have been turned down. This incurs certain normalization factors that one needs to keep track of.

Further we can use the locality principle (See Fig. 7.7) to give us Fig. 10.7

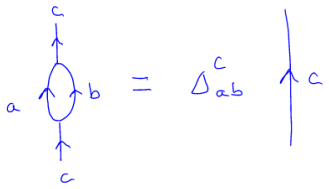


Figure 10.7: Removal of a bubble gives a factor, which we call $\Delta_{ab}^{c} \neq 0$.

We can then use these two identities to directly fuse the loop of a with the loop of b incurring a factor of $\kappa_{ab}^{c}\Delta_{ab}^{c}$ as shown in Fig. 10.8

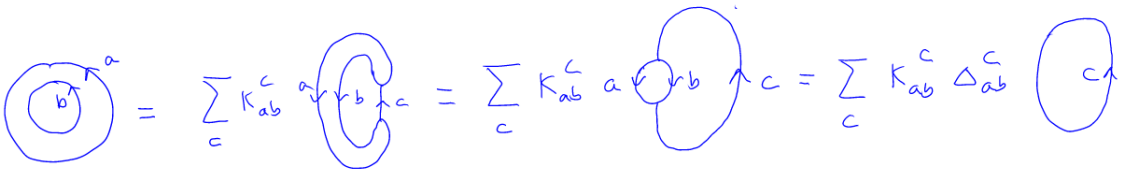


Figure 10.8: We have applied first the result of Fig. 10.6 then Fig. 10.7. Note that if a and b cannot fuse to c then that term is zero in the sum.

However, we can also apply the same reasoning to split the loops into multiple bubbles as shown in Fig. 10.9.

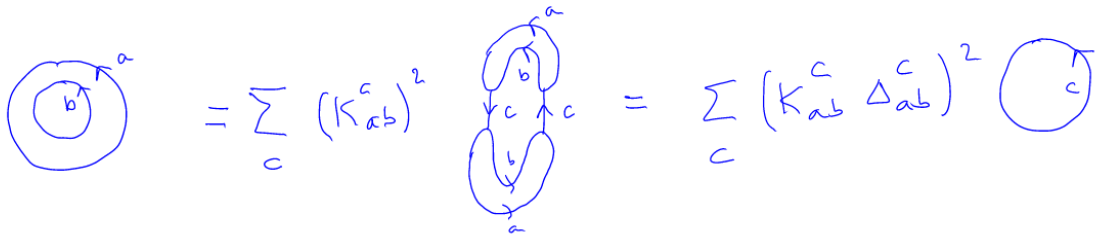


Figure 10.9: Applying the result of Fig. 10.6 twice then Fig. 10.7 twice.

From these two results we can immediately conclude that $\kappa_{ab}^c \Delta_{ab}^c = 0$ or 1. Since both of these factors are nonzero when a and b can fuse to c, and are zero when they cannot, we can write $\kappa_{ab}^c \Delta_{ab}^c = N_{ab}^c$ (assuming no $N_{ab}^c > 1$)[4]. This then proves our Lemma.

Once it is established that the factor $\kappa_{ab}^c \Delta_{ab}^c = N_{ab}^c$ then this can be also used to directly prove the identity in Fig. 10.4.

[4]In cases where $N_{ab}^c > 1$ we would have had to keep track of an additional index μ at the a, b, c vertex. However, this index is also conserved around the loop meaning that the sum eventually becomes $\sum_{c,\mu}$ which will then generate a factor of N_{ab}^c as desired.

Chapter 11

Quantum Error Correction and The Toric Code

We now change subjects a bit towards quantum error correction and the toric code. While initially the ideas may seem somewhat different from what we have been discussing, we will see that it is extremely closely related and brings us to an extremely important application of many of the ideas we have been discussing.

11.1 Classical Versus Quantum Information

11.1.1 Memories[1]

Classical Memory

The unit of classical information is a bit — classical two state system which can take the values 0 or 1. A memory with N bits can be in any one of 2^N states — each state corresponding to a particular bit-string, such as 011100111.

Quantum Memory

The unit of quantum information is the quantum bit or qubit[2] which is a quantum two state system – i.e. a 2-dimensional complex Hilbert space spanned by vectors $|0\rangle$ and $|1\rangle$. A qubit can be in any state

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$$

with arbitary complex prefactor α, β (where we normalize wavefunctions so $|\alpha|^2 + |\beta|^2 = 1$.

A quantum memory with N qubits is a vector within the 2^N dimensional complex Hilbert space. So for example, with 2 qubits the general state of a system is specified by four complex parameters

$$|\psi\rangle = \alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle \tag{11.1}$$

with the normalization condition $|\alpha|^2 + |\beta|^2 + |\gamma|^2 + |\delta|^2 = 1$. So to specify the state of a quantum memory with 2 bits, you have to specify four complex parameters, rather than, in the classical case just stating which of the four states the system is in!

11.1.2 Errors

An error is some process which accidentally changes the state of the memory away from the intended state. Often we take as an error model the case where only one bit or one qubit is effected at a time (a "minimal" error) although more complicated errors can occur.

[1] All alone in the moonlight!

[2] Sometimes q-bit, but never cubit.

Classical Error Correction There is a simple way to correct small errors for a classical memory. Instead of storing a single bit 0 or 1, instead store multiple copies of the bit (say, three copies). So we use three physical bits to store one "logical" bit of information. Our memory should either be in

logical bit	physical bits
0	000
1	111

Table 11.1: Three bit repetition code. Stores a single logical bit of information using three physical bits.

the state 000 or 111 — we call these two possibilities the *code space*. If we detect the system being in any other state of the three bits (i.e., not in the code space) we know an error has occured. If an error does occurs on one of the physical bits (i.e,, if one of the bits is accidentally fliped) we can easily find it, because it would leave our memory with not all of the physical bits being the same. For example, if our system starts as 000, an error introduced on the second bit would leave it in the form 010. But then, by just using a majority-rule correction system, it is easy to figure out what happened and flip the mistaken bit back. So our error correction protocol would be to continuously compare all three bits, if they don't match, switch the one back which would bring them back to matching. Assuming errors are rare enough (and only occur on one bit at a time) this scheme is an effective way to prevent errors. For added protection one can use more redundant bits, such as 5 bits or 7 bits.

One might think the same sort of approach would work in the quantum world: make several copies of the qubit you want to protect, and then compare them to see if one has changed. Unfortunately, there are two big problems with this. The first is the so-called no-cloning theorem — it is not possible to make a perfect clone of a qubit. The second reason is that measuring a state inevitably changes it.

Quantum No Cloning Theorem: (Zurek et al 1982). The result is such a straightforward result of quantum mechanics some people have argued whether it deserves to be called a theorem. The statement of the "theorem" is as follows:

> **Theorem:** Given a qubit in an arbitrary unknown state $|\phi_1\rangle$ and another qubit in an initial state $|\phi_2\rangle$, there does not exist any unitary operator U (i.e., any quantum mecahnical evolution) such that
>
> $$U(|\phi_1\rangle \otimes |\phi_2\rangle) = e^{i\chi}|\phi_1\rangle \otimes |\phi_1\rangle$$
>
> for all possible input $|\phi_1\rangle$.

The point here is that we do not have a way to copy $|\phi_1\rangle$ into the auxiliary qubit $|\phi_2\rangle$.

Proof of Theorem: Suppose we have two states $|0\rangle$ and $|1\rangle$ which are properly copied.

$$U(|0\rangle \otimes |\phi_2\rangle) = e^{i\chi}|0\rangle \otimes |0\rangle$$
$$U(|1\rangle \otimes |\phi_2\rangle) = e^{i\chi}|1\rangle \otimes |1\rangle$$

Now quantum mechanical operators are linear so we can try applying this operator to the linear superposition $\alpha|0\rangle + \beta|1\rangle$ and we must get

$$U([\alpha|0\rangle + \beta|1\rangle] \otimes |\phi_2\rangle) = e^{i\chi}(\alpha|0\rangle \otimes |0\rangle + \beta|1\rangle \otimes |1\rangle)$$

but this is now *not* what a putative cloning device must give. Instead it should have given the outcome

$$e^{i\chi}[\alpha|0\rangle + \beta|1\rangle] \otimes [\alpha|0\rangle + \beta|1\rangle]$$

Thus no cloning device is consistent with the linearity inherent in quantum mechanical evolution.

11.2 The Toric Code

Perhaps the most surprising thing about quantum error correction is that it is possible at all! This was discovered by Peter Shor in 1995 (and shortly thereafter by Andrew Steane). We will describe the Toric code approach to error correction which is potentially the conceptually most simple error correction scheme, as well as being very possibly the most practical to implement in real systems[3]!

As with so many great ideas in this field, the Toric code was invented by Kitaev (Kitaev 1997).

11.2.1 Toric Code Hilbert Space

We imagine an N_x by N_y square lattice with spins on each edge, where the edges of the lattice are made periodic hence forming a torus (hence the name). The total number of spins is $N = 2N_xN_y$ and correspondingly the dimension of the Hilbert space is 2^N.

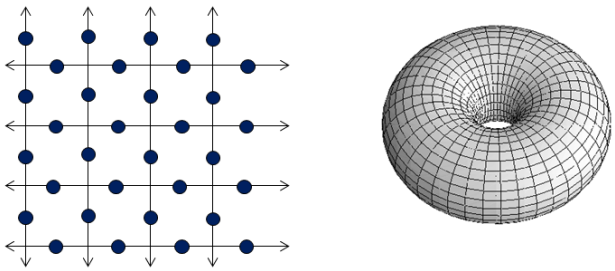


Figure 11.1: The Hilbert space of the toric code — an N_x by N_y square lattice with spins (dots) on each edge wrapped up to make it periodic in both directions — i.e., a torus. Hence the name. There are 32 spins in this picture so the Hilbert space has dimension 2^{32}.

We will work with a basis in our Hilbert space of up and down spins[4]. A convenient notation is then to color in the edges containing down spins but leave uncolored the edges with up spins. See Fig. 11.2.

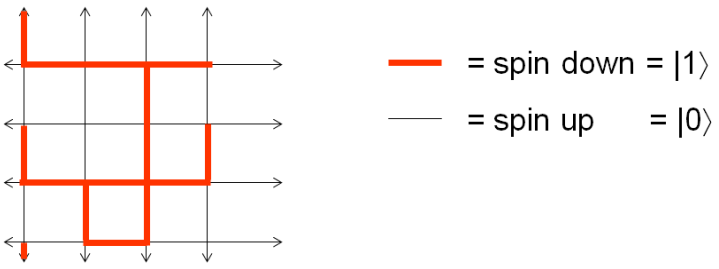


Figure 11.2: A particular basis state of the Hilbert space, working in the up-dpwn basis (z-eigenstates). Here we denote down spins by thick (red) lines. And up spins are denoted by not coloring in the edges.

[3]The statement that it is the most practical is based on the fact that the so-called surface codes (which is essentially the toric code) has the highest known error threshold — meaning you can successfully correct even highly faulty qubits with this technique compared to other techniques which require your qubits to be much closer to perfect to begin with.

[4]Caution: In the literature about half of the world uses the up-down or z-eigenstates as a basis, and half of the world uses the x-eigenstates as a basis.

Note that it is not crucial that we are working with a square lattice, or that we are even working on a torus (although it is crucial that the surface has noncontractable loops). We could work with other types of lattices — the honeycomb will be useful later — or triangular lattice could be used. In fact even irregular lattices (which are not really lattices, since they are irregular, and should be called 'graphs') can be used. However it is a lot easier to continue the discussion on this simple square-lattice-torus geometry.

11.2.2 Vertex and Plaquette Operators

Let us now define some simple operators on this Hilbert space.

First, given a vertex α which consists of four incident edges $i \in \alpha$, we define the vertex operator

$$V_\alpha = \prod_{i \in vertex\ \alpha} \sigma_i^z$$

This operator simply counts the parity of the number of down spins (number of colored edges) incident on the vertex. It returns $+1$ if there are an even number of incident down spins at that vertex and returns -1 if there are an odd number. (And in either case, as is obvious $V_\alpha^2 = 1$). This is depicted graphically in Fig. 11.3.

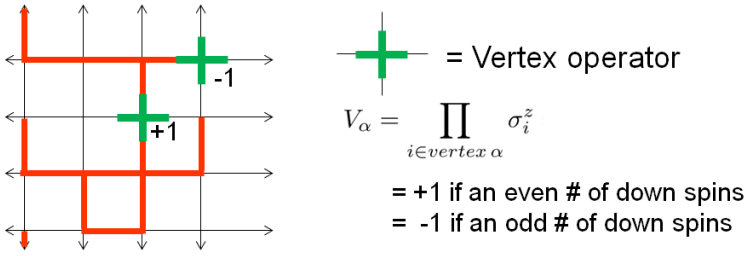


Figure 11.3: The vertex operator returns $+1$ if there are an even number of incident down spins at that vertex and returns -1 if there are an odd number.

There are a total of $N_x N_y$ vertex operators.

Note that it is possible (and useful) to define a corresponding projection operator

$$\tilde{V}_\alpha = \frac{1}{2}(1 - V_\alpha) \tag{11.2}$$

which has eigenvalues 0 for an even number of incident down spins or 1 for an odd number. This is a projection operator because $\tilde{V}_\alpha = \tilde{V}_\alpha^2$.

We now define a slightly more complicated operator known as the plaquette operator. Given a plaquette β which contains four edges in a square $i \in \beta$ we define

$$P_\beta = \prod_{i \in plaquette\ \beta} \sigma_i^x$$

which flips the state of the spins on all of the edges of the plaquette as depicted in Fig. 11.4. There are a total of $N_x N_y$ plaquette operators.

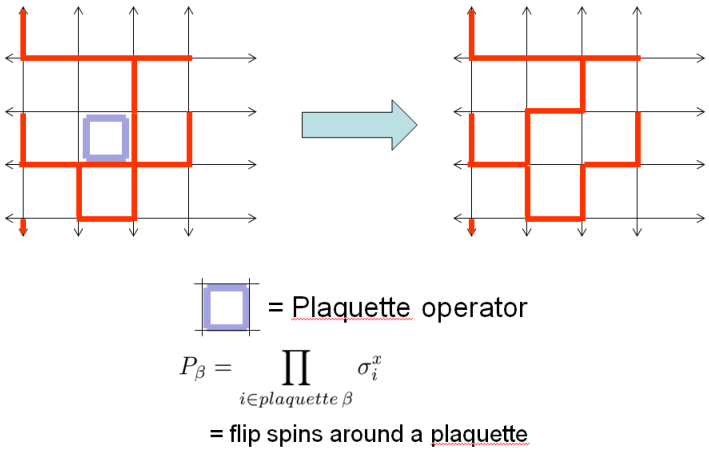


Figure 11.4: The plaquette operator flips the state of the spin on the four edges of a plaquette.

As with the vertex operator $P_\beta^2 = 1$ meaning P_β has eigenvalues $+1$ and -1. As with the vertex operator, we can define a projector

$$\tilde{P}_\beta = \frac{1}{2}(1 - P_\beta) \tag{11.3}$$

which satisfies $P_\beta^2 = P_\beta$.

It is a bit more difficult to describe what these eigenstates of the plaquette operators are. In the basis we are using, the spin-up/spin-down basis corresponding to uncolored and colored edges, the P_β operator is off-diagonal — it flips spins around a plquette. As such, the 0 eigenstate of $\tilde{P}_\beta$ operator (i.e, the 1 eigenstate of P_β) is obtained by adding the state of a plaquette to the flipped state of the plaquette as shown in Fig. 11.5. The orthogonal superposition (adding the two states with a - sign) will give the other eigenstate.

Figure 11.5: A linear superposition of a flipped and unflipped plaquette is a $+1$ eigenstate of P_β or equivalently a 0 eigenstate of $\tilde{P}_\beta$. The -1 eigenstate is given by the orthogonal superposition, i..e, the superposition with a - sign between the two terms.

Operators Commute

I claim all of the plaquette operators and all of the vertex operators commute with each other. It is obvious that

$$[V_\alpha, V_{\alpha'}] = 0$$

since V_α's are only made of σ_z operators and all of these commute with each other. Similarly

$$[P_\beta, P_{\beta'}] = 0$$

since P_β's are made only of σ_x operators and all of these commute with each other.

The nontrivial statement is that

$$[V_\alpha, P_\beta] = 0$$

for all α and β. The obvious case is when V_α and P_β do not share any edges — then the two operators obviously commute. When they do share edges, geometrically they must share exactly two edges, in which case the commutation between each shared σ_i^x and σ_i^z accumulates a minus sign, and there are exactly two shared edges so that the net sign accumulated is $+1$ meaning that the two oprators commute.

Is the set of operators complete?

We have $N_x N_y$ vertex operators and $N_x N_y$ plaquette operators — all of these operators commute, and each of these operators has 2 eigenvalues. This appears to match the fact that there are $2N_xN_y$ spins in the system. So is our set of V and P operators a complete set of operators on this Hilbert space? (I.e., is it true that describing the eigenvalue of each of these operators must determine a unique state of the Hilbert space?)

It turns out that the V and P operators do not quite form a complete set of operators on the Hilbert space. The reason for this is that there are two constraints on these operators

$$\begin{aligned} \prod_\alpha V_\alpha &= 1 \\ \prod_\beta P_\beta &= 1 \end{aligned}$$

To see that these are true, note that each edge occurs in exactly two operators V_α. Thus when we multiply all the V_α's together, each σ_i^z occurs exactly twice, and $(\sigma_i^z)^2 = 1$. Thus the product of all the V_α's is the identity. The argument is precisely the same for multiplying together all of the P_β's.

Thus we can freely specify the eigenvalues of $(N_xN_y - 1)$ operators V_α, but then the value of the one remaining V_α is then fixed by the values chosen for the other $(N_xN_y - 1)$ of them. Similarly with the P_β's. So specifying the eigenvalues of these commuting operators specifies only $2(N_xN_y - 1)$ degrees of freedom, and since we started with $2N_xN_y$ spins, we still have 2 degrees of freedom remaining. These two degrees of freedom are going to be two error protected qubits in this scheme for building an quantum error correcting code.

Note that this result, of having two degrees of freedom that remain unspecified by the plaquette and vertex opertaors, is not unique to having used a square lattice (we can use triangular lattice, honeycomb, or even irregular grids), but depends only on having used a torus. If we use a g-handled torus we will have $2g$ degrees of freedom (i.e., $2g$ qubits) remaining. To see this we use the famous Euler characteristic. For any decompositon of an orientable 2-manifold into a grid, we have the formula

$$2 - 2g = (\text{Number of Vertices}) - (\text{Number of Edges}) + (\text{Number of Faces})$$

where g is the number of handles on the manifold. Since there is one spin on each edge we have

$$\begin{aligned} &\text{Number of Vertex Operators} + \text{Number of Plaquette Operators} - 2 + 2g \\ =&\text{Number of Spins} \end{aligned}$$

We can read this as follows. The right hand side is the total number of degrees of freedom. On the left we can specify all the eigenvalues of the vertex and plaquette operators, then there are 2 constraints, so subtract two, and this leaves us with $2g$ unspecified degrees of freedom.

11.2.3 Building the code space

We are going to state two rules for constructing our code. We are imagining here that we have a great deal of control over the spins (the microscopic qubits) making up our system and we can impose these rules by fiat.

Rule 1: Specify that $V_\alpha = 1$ for every vertex (or equivalently $\tilde{V}_\alpha = 0$.).

This assures that there are an even number of down spins (red lines) incident on every vertex. It is easy to see that this can be interpreted as a constraint that one must consider only loop

configurations of these red lines. There can be no ends of lines, and no branching of lines. See, for example, fig. 11.6

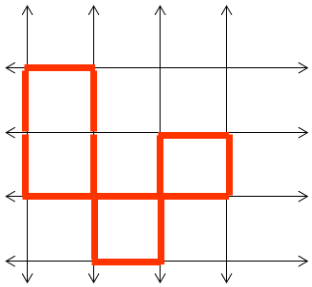


Figure 11.6: A loop configuration consistent with the constraint that $V_\alpha = 1$ on every vertex. There must be an even number of red lines incident on every vertex.

The idea of an error correcting code is that once we construct our code, we will have some way to check that this Rule 1 is satisfied and if it is not satisfied we should have some way to fix it without destroying our encoded quantum information.

Rule 2: Specify that $P_\beta = 1$ for every plaquette (or equivalently $\tilde{P}_\beta = 0$.).

As mentioned above in Fig. 11.5 this assures that every plaquette is in an equal superpositon of flipped and unflipped states with a plus sign between the two pieces. Note in particular that, because the P_β and V_α operators commute, the action of flipping a plaquette will not ruin the fact that Rule 1 is satisfied (that is, that we are in a loop configuration).

The quantities V_α and P_β are known as the *stabiizers* of the code — they are meant to stay constant and are checked for any errors which are indicated by the fact that their value has changed.

We thus have the following prescription for constructing a wavefucntion that satisfies both Rule 1 and Rule 2: First start in any state of spins up and spins down which satisfies rule 1, i.e., is a loop configuration. Then add to this in a superposition every configuration that can be obtained by flipping plaquettes. We thus have

$$|\psi\rangle = \sum_{\substack{\text{all loop configs that can be obtained by flipping plaquettes from a reference loop config}}} |\text{loop config}\rangle \tag{11.4}$$

By adding up all such configurations, we assure that every plaquette is in the correct superpositon of flipped and upflipped and we satisfy Rule 2.

The key queston is whether one can obtain all loop configurations by starting in a referecnce configuratation and flipping plaquettes. The answer is that you cannot: Flipping plaquettes never changes the *parity* of the number of loops running around the handle. To see this, try making a cut around a handle of the torus, as shown in Fig. 11.7. If one flips a plaquette (blue in the fig) along this cut (green inn the fig), it does not change the number of red bonds that the cut goes through.

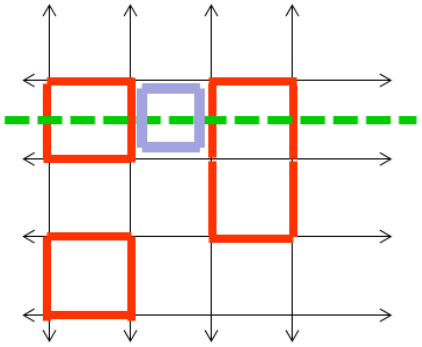


Figure 11.7: Making a cut around one of the handles of torus, one can see that flipping a plaqutte, such as the blue one, does not change the parity of the number of red bonds cutting the green line. Further, it does not matter where (at which y-coordinate) the green cut is made, the number of red bonds it cuts is always even.

Thus there are four independent wavefunctions of the form of Eq. 11.4, which are different in whether the reference configuration has an even or an odd number of red bonds going around each handle. All of these states satisfy the constraints rules that all $V_\alpha = 1$ and all $P_\beta = 1$. We will call these states

$$|\psi_{ee}\rangle \quad |\psi_{eo}\rangle \quad |\psi_{oe}\rangle \quad |\psi_{oo}\rangle$$

where e and o stand for an even or an odd number of red lines going around a given handle. So for example, we have

$$|\psi_{ee}\rangle = \sum_{\substack{\text{all loop configs that have} \\ \text{an even number of red} \\ \text{bonds around both handles}}} |\text{loop config}\rangle$$

Or graphically, we have Fig. 11.8

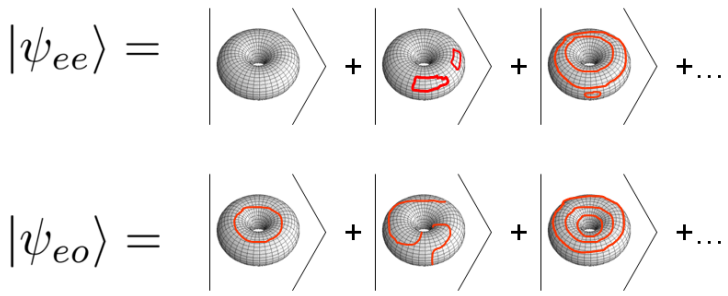


Figure 11.8: Graphical depiction of $|\psi_{ee}\rangle$ which has an even number of strings running around each handle, and $|\psi_{eo}\rangle$ which is even around the first handle odd around the second.

The most general wavefunction we can write that satisfies the two above rules, that all $V_\alpha = 1$ and all $P_\beta = 1$ is thus of the form

$$|\psi\rangle = A_{ee}|\psi_{ee}\rangle + A_{eo}|\psi_{eo}\rangle + A_{oe}|\psi_{oe}\rangle + A_{oo}|\psi_{oo}\rangle \tag{11.5}$$

for arbitrary coefficients $A_{ee}, A_{eo}, A_{oe}, A_{oo}$. It is these coefficients which are the two qubits of quantum information that we are trying to protect with this coding scheme (exactly like Eq. 11.1). We will refer to wavefunctions of the form of Eq. 11.5 as the "code-space". We refer to these two bits as being the "logical" qubits – the information we are trying to protect. The underlying spins on the lattice that make up the code are sometimes called the "physical" qubits.

Note that in order to turn the $|\psi_{ee}\rangle$ wavefunction into the $|\psi_{eo}\rangle$ we need to insert a single loop around a handle — this involves flipping an entire row of spins at once. If one were to try to flip only some of these spins, we would have an incomplete loop — or an endpoint – which violates the rule that $V_\alpha = 1$ for all vertex sites — i.e, not in the code-space. It is this fact that allows us to test for errors and correct them efficiently, as we shall see.

11.3 Errors and Error Correction

Let us now turn to study possible errors in more detail. What does an error look like in this system? Imagine a demon arrives and, unbeknownst to us, applies an operator to one of the spins in the system.

11.3.1 σ_x errors

Let us first consider the case where that operator happens to be a σ^x on bond i. This operator commutes with all the plaquette operators P_β but anticommutes with the vertex operators V_α which intersect that bond. This means, if we start in the code space (all $V_\alpha = +1$), and apply this error operator σ_i^x, we then end up in a situation where the the two vertices attached to the bond i are now in the wrong eigenstate $V_\alpha = -1$. To see this more clearly starting in the original state $|\psi\rangle$ we have

$$V_\alpha|\psi\rangle = |\psi\rangle$$

meaning we start in the $+1$ eigenstate, now apply the error operator σ_i^x to both sides

$$\sigma_i^x|\psi\rangle = \sigma_x V_\alpha|\psi\rangle = -V_\alpha\sigma_i^x|\psi\rangle$$

or

$$V_\alpha[\sigma_i^x|\psi\rangle] = -[\sigma_i^x|\psi\rangle]$$

showing we end up in the -1 eigenstate of the vertex operator.

To show these errors graphically we will no longer draw the up and down spins (the red bonds) but instead we just draw the σ_x operator as a blue line, and the vertices which are in the -1 eigenstate as a red X as shown in Fig. 11.9.

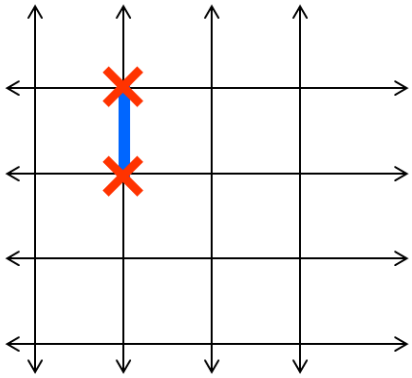


Figure 11.9: A σ^x operator applied to the bond creates two vertices in the $V_\alpha = -1$ eigenstate.

So it is clear what our error correction protocol must do. It must frequently measure the state of the V_α operators, and if it finds a pair in the $V = -1$ state, we know that a σ^x has been applied on the intervening bond. Once we have identified the error it is easy to correct it by applying σ_x on the same bond, thus returning the system to its original state and to the code space.

Now suppose that the demon is very fast and manages to make several such errors very quickly. If these errors are well separated from each other, we will easily find multiple pairs of vertices in the $V = -1$ state, with the pair separated from each other by one bond distance. These can similarly be caught by our correction scheme and repaired, returing us to the code space again.

However, it could be the case that two errors are on bonds that share a vertex , as shown on the left of Fig. 11.10, the vertex that is shared gets hit by σ^x twice and is thus in the $V = +1$ state. Only the two vertices at the end of the "string" are in the $V = -1$ state and are then detectable as errors.

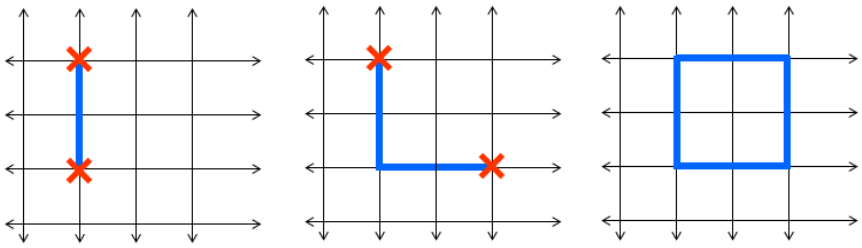


Figure 11.10: **Left:** When two σ^x errors are made on bonds that share a vertex, the shared vertex is hit with σ_i^x twice, and thus becomes $V = +1$ again. Only the two vertices at the end of the "string" are in the $V = -1$ state. **Middle:** A longer string of errors. Note that we can only measure the endpoints of the string, not where the errors were made, so we cannot tell if the error string goes down two steps then two steps to the right, or if goes two steps to the right then down two steps. **Right** If we detect the errors as in the middle panel and we try to correct it by dragging the errors back together, but we choose the incorrect path for the string, we end up making a closed loop of σ_x operators – which acts as the identity on the code space, so we still successfully correct the error!

Nonetheless, the error correction scheme is still fairly straightforward. One frequently checks the state of all the vertices and when $V = -1$ is found, one tries to find the closest other error to pair it with – and then apply σ_x operators to correct these errors (you can think of this as dragging the errors back together and annihilating them with each other again).

It is important to realize that we cannot see the error operators (which we have drawn as a blue string) themselves by making measurements on the system – we can only detect the endpoints of string, the vertices where $V = -1$. For example, in the middle panel of figure 11.10 we cannot tell if the error string goes down two step and then to the right, or if it goes to the right one step and then down two steps. We only know where the endpoints of the string are.

Now if we detect the two errors in the middle panel of Fig. 11.10, we may try to correct these errors by guessing where the blue string is and applying σ_x along this path to bring the endpoints back together and reannihilate them. However, it is possible we guess incorrectly as shown in the right panel of Fig. 11.10. In this case we will have ended up producing a closed loop of σ_x operators applied to the original state. However, a product of σ^x operators around a closed loop is precisely equal to the product of the plaquette operators P_β enclosed in the loop. Since the code space is defined such that all of hte plaquettes operators are in the $+1$ eigenstate, this loop of σ^x acts as the identity on the code space, and we still successfull correct the error.

On the other hand, if a loop of errors occurs which extends around a handle , and the $V = -1$ errors annihilate again (think of this as dragging the error all the way around the handle and re-annihilating it again) then, although we return to the code-space (there are no $V = -1$ verteices) we have changed the parity of the number of down spins around a handle thus scrambling the quantum information and make an error in the logical bits. In fact what we get in this case is the transform

that switches the even and odd sectors around one handle :

$$A_{ee}|\psi_{ee}\rangle + A_{eo}|\psi_{eo}\rangle + A_{oe}|\psi_{oe}\rangle + A_{oo}|\psi_{oo}\rangle \longrightarrow$$
$$A_{oe}|\psi_{ee}\rangle + A_{oo}|\psi_{eo}\rangle + A_{ee}|\psi_{oe}\rangle + A_{eo}|\psi_{oo}\rangle$$

However, the general idea of the toric code is that by having a very large torus, it requires a very large number of errors to make this loop around the handle and actually scramble the quantum information (the logical qubits). If we are continuously checking for $V = -1$ errors we can presumably correct these errors before a logical error can arise.

11.3.2 σ_z errors

We can also consider what happens if the error is not a σ^x operator applied to the system, but rather a σ^z operator. Much of the argument in this case is similar to that above.

Since the σ^z operator on an edge anticommutes with the two neighboring plaquettes P_β which share that edge, the resulting state will have $P_\beta = -1$ for these two plaquettes as shown on the left of Fig. 11.11. Recall that this eigenstate of the plaquette operator is a superposition of the flipped and unflipped plquettes similar to that shown in Fig. 11.5 but with a minus sign between the two terms.

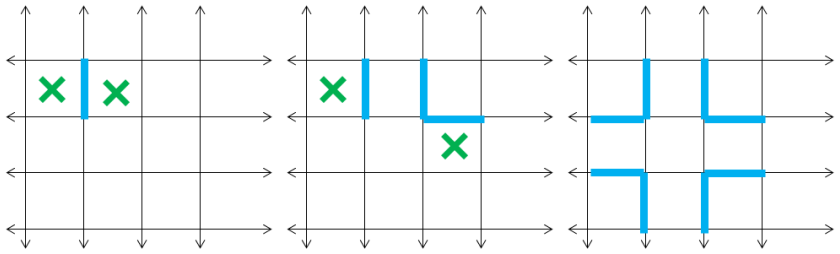


Figure 11.11: **Left:** When a σ^z error is applied to a bond, the plaquettes on either side end up in the $P = -1$ state **Middle:** A string of several σ^z errors. **Right** A closed loop of σ^z errors. This is equal to the product of all of the enclosed V_α operators. In the code space, this is equal to $+1$.

Analogous to the above discussion, our σ^z error correction protocol should frequently check for pairs of neighboring plaquettes where $P_\beta = -1$ and if these are found the protocol should correct the error by applying σ^z to the intervening edge. As above, if several σ^z errors are created, they can form a string, as shown as blue bonds in the middle of Fig. 11.11. As above, one is not able to actually detect the string, but can only see the endpoints as plaquettes where $P = -1$. Analogous to the above case, if from errors, or from an attempt to correct errors, the σ^z error string forms a closed loop as in the right of Fig. 11.11, this loop of σ^z operators is equal to the product of the enclosed V_α operators. Since within the code space, $V_\alpha = 1$, a closed loop returns the system its original state. Another way of seeing this is to think in terms of the red loops of down spins discussed above. The σ_z operators register -1 each time they intersect a red loop. On the other hand the red loops must be closed so the number of intersections between a red loop and a dlow3e loop of the blue σ^z error string in the figure must be even (since a red loop going into the region surrounded by the string must also come out), thus forcing the product of the blue σ^z operators to have a value of 1.

On the other hand, if the loop of σ^z operators goes all the way around the handle, it then scrambles the logical qubits. In particular, one can see that if there is a string of σ_z going all the way around a handle as shown as the blue bonds in Fig. 11.12, this operator then counts the parity of the number of red bonds going around the dual handle, as shown in the figure.

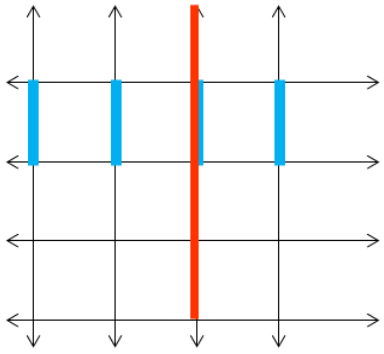


Figure 11.12: If a string of σ^z goes around a handle, it measures the parity of the number of red strings going around the dual handle.

Thus, applying the string of σ^z operators around the handle makes the transformation

$$A_{ee}|\psi_{ee}\rangle + A_{eo}|\psi_{eo}\rangle + A_{oe}|\psi_{oe}\rangle + A_{oo}|\psi_{oo}\rangle \longrightarrow$$
$$A_{ee}|\psi_{ee}\rangle + A_{eo}|\psi_{eo}\rangle - A_{oe}|\psi_{oe}\rangle - A_{oo}|\psi_{oo}\rangle$$

11.3.3 σ^y errors

A basis for a complete set of operators applied to a single spin is given by σ^x, σ^y, and σ^z (as well as the identity). We have discussed errors created by σ^x and σ^z, but what about σ^y. Here we simply use the fact that

$$\sigma^y = i\sigma^x\sigma^z$$

So if we have an error correction protocol that removes both σ^x and σ^z errors, being that the two procedures don't interfere with each other, we will automatically correct σ^y errors in the process!

11.3.4 More Comments on Errors

(1) A key point to take away here is that the *only* process which can cause logical erorrs is if an error string goes all the way around one of the handles. Further (and this is a related statement) the only operator that can distinguish the different elements of the code space from each other are string operators that go all the way around the handles. The latter (related) statement is qutie necessary, since being able to distinguish the different wavefunctions from each other is equivalent to causing an error since it amounts to a measuremtn of the logical bits.

(2) As mentioned above, the toric code as method of storing quantum information is considered the "best" quantum error correcting code. We define the quality of a code as follows: We define a time unit as the amount of time it takes us to make a measurement of a quantity such as V_α or P_β. Then we assume there is some rate of errors being introduced to they underlying physical bits (the spins) per time unit. Given these parameters, the toric code is able to reliably correct the largest possible error rate per time unit of any known quantum error correcting code.

(3) While we have introduced the toric code on a torus (hence the name) so that it stores 2 logical qubits of information, as mentioned above, if we go to a higher genus surface (either a closed manifold with handles, or a surface with holes cut in it) we can store $2g$ qubits where g is the genus of the surface.

11.4 Toric Code as Topological Matter

We have introduced the toric code as a way to store quantum information — being stabilized by an error correction protocol that actively checks the value of the vertex and plaquette operators. However, it is quite easy to convert this story to a a realization of **topologically ordered quantum matter** — a physical system that is described at low temperature and long wavelength by a topological quantum field theory. In this case the physical system will be stabilized by the existence of an energy gap to excitations and the fact that our system will be kept at low temperature.

To recast the toric code as topologically ordered matter, we simply write a Hamiltonian which is a sum of commuting operators

$$H = -\sum_{vertices\,\alpha} V_\alpha - \sum_{plaquettes\,\beta} P_\beta \tag{11.6}$$

Here we have set the energy unit to unity. The Hamiltonian is made of a sum of commuting projectors with eigenvalues ± 1 so the ground state space is described by simply setting all of the $V_\alpha = 1$ and $P_\beta = 1$. I.e., the ground state space is exactly the code space. There will be a four-fold degenerate ground state corresponding to the four orthogonal wavefunctions in the code space. If $V_\alpha = -1$ or $P_\beta = -1$ this corresponds to a particle excited out of the ground state.

It is sometimes more convenient to work with the projectors $\tilde{V}_\alpha$ and $\tilde{P}_\beta$ defined by Eqs. 11.2 and 11.3. Writing

$$\tilde{H} = \sum_{vertices\,\alpha} \tilde{V}_\alpha + \sum_{plaquettes\,\beta} \tilde{P}_\beta \tag{11.7}$$

which differs from Eq. 11.6 only be a factor of 2 and an overall constant. The advantage of $\tilde{H}$ is that it is a sum of commuting projection operators. This is often convenient because it means that the ground state has energy 0 and each excitatation has unit energy.

11.4.1 Excitations

The types of particle-excitations we can have are given as follows:

(1) We can have a vertex where $V_\alpha = -1$ instead of $V_\alpha = +1$. We call this an "electric particle" which we write as e.

(2) We can have a plaquette where $P_\beta = -1$ instead of $P_\beta = +1$. We call this a "magnetic particle" which we write as m.

The nomenclature for these particles due to a relationship with lattice gauge theories which we will discuss below.

Since vertex defects e's are produced in pairs, and can be brought back together and annihilted in pairs, we know we must have

$$e \times e = I$$

Similarly since plaquette defects m are produced in pairs, and can be brought back together and annihilated in pairs we must also have

$$m \times m = I$$

We might then wonder what happens if we bring together a vertex and a plaquette defect. They certainly do not annihilate, so we define another particle type, called f, which is the fusion of the two

$$e \times m = f$$

We then have

$$f \times f = I$$

which we can see by associativity and commutativity

$$f \times f = (e \times m) \times (e \times m) = (e \times e) \times (m \times m) = I \times I = I$$

These are the only particle types there are. Note that they form a closed set under the fusion rules. There are no non-abelian fusions here so we assume we have an abelian model of some sort.

Note that there are exactly four particle types (including the identity), and there are exactly four ground states!

The full fusion relations are given by

$\times$	I	e	m	f
I	I	e	m	f
e	e	I	f	m
m	m	f	I	e
f	f	m	e	I

11.4.2 Braiding Properties

e is a boson

Let us first consider the e particles. These are both created and moved around by applying σ_x operators. All of the σ_x operators commute with each other, so there should be no difference in what order we create, move, and annihilate the e particles. This necessarily implies that the e particles are bosons. There are several "experiments" we can do to sow this fact. For example, we can create a pair of e's move one around in a circle and reannihilate, then compare this to what happens if we put another e inside the loop before the experiment. We see that the presence of another e inside the loop does not alter the phase of moving the e around in a circle[5].

m is a boson

Entirely analogously we can argue that m is also a boson. m is both created and moved by the σ^z operator and all of these operators commute with each other. The exact same argument (here without detail) shows us that m must be a boson.

[5]The experiment just described, while quite clear only tells us that e is either a boson or a fermion (since a fermion taken in a loop all the way around another fermion also accumulates no phase since it is equivalent to two exchanges).

To determine the phase of an exchange, we are going to attempt to do a twist in a world line as in Fig. 2.5 or 8.1. Considering Fig. 11.13

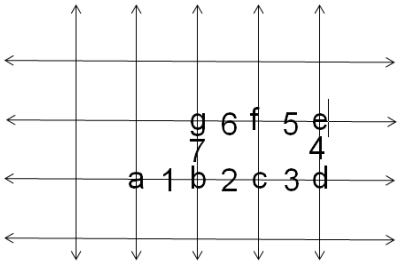


Figure 11.13: Vertices are labeled with letters and bonds are labeled with numbers.

Now suppose there is initially an e particle at position a. One experiment we can do is to apply (reading right to left) $\sigma_1^x\sigma_7^x\sigma_6^x\sigma_5^x\sigma_4^x\sigma_3^x\sigma_2^x\sigma_1^x$. This just moves the particle starting at a around in a loop (reading right to left $abgfedcba$) and brings it back to the original position. We can compare this to the following operations $\sigma_1^x\sigma_2^x\sigma_1^x\sigma_7^x\sigma_6^x\sigma_5^x\sigma_4^x\sigma_3^x$. This instead creates a pair of e particles at positons c and d, moves the particle at d in a loop ($bgfe$) around c and anihilates it with the particle at a, then finally moves the particle from e to replace the particle initially at a. This process is precisely the twist factor process from Fig. 2.5 or 8.1. However, since the σ_x operators all commute, it must also be equal to the previously described process which just moves one particle around in a loop without introducing any twist. Hence we conclude that the e particle is a boson.

Braiding e and m

Here is where it gets interesting. Suppose we create an e particle and move it around in a circle then reannihilate. This is exactly the process shown in the right panel of Fig. 11.10 and is the product of a string of σ^x operators. Recall that the reason this process does not accumulate a phase is because the string of σ^x operators around the loop is equivalent to the product of the P_β plaquette operators enclosed — and in the ground state, the P_β operators are in the $+1$ state. However, if there is one m particle inside the loop, this means that one of the P_β operators is actually in the -1 state. In this case the phase of taking the e particle around in a loop is actually -1. So there is a phase of -1 for taking e around m.

We can check that it is precisely equivalent if we take an m particle around an e. Taking an m around in a loop is the process shown on the right of Fig. 11.11 and is the product of a string of σ^z operators. Recall that the reason this process does not accumulate a phase is because the string of σ^z operators around the loop is equivalent to the product of the V_α vertex operators enclosed — and in the ground state, the V_α operators are in the $+1$ state. However, if there is one e particle inside the loop, this means that one of the V_α operators is actually in the -1 state. In this case the phase of taking the m particle around in a loop is actually -1. So there is a phase of -1 for taking m around e.

Properties of f, the fermion

Since f is made up of an m bound to an e, it is easy to see that taking e around f accumulates a phase of -1 and taking m around f also accumulates a phsae of -1. More interesting is the properties of a single f. We claim that f is a fermion. The easiest way to see this is to check its phase under a twist as shown in Fig. 11.14

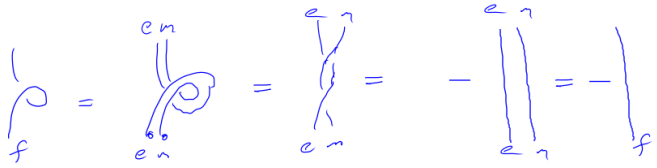


Figure 11.14: The $f = e \times m$ particle is a fermion, since e braiding around m gives a -1 sign.

Note that taking f all the way around f will result in a net $+$ sign.

11.4.3 Modular S-matrix

We can summarize these findings with a modular S_{ij} matrix, which lists the braiding result obtained by taking particle i around particle j as shown in Fig.6.12. Listing the particles in the order I, e, m, f we can write S as in

$$S = \frac{1}{\mathcal{D}} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

where unitarity fixes the total quantum dimension $\mathcal{D} = 2$.

11.4.4 Flux Binding Description

We can describe the physics of the toric code phase in a flux binding description somewhat analogous to Chern-Simons theory. Here let us define

electric particle = e = particle bound to 1 unit of electric charge
magnetic particle = m = particle bound to π units of magnetic flux
fermion = f = particle bound to 1 unit of electric charge and π units of magnetic flux

It is easy to see that this charge and flux will correctly give the +1 and -1 phases accumulated from braiding particles.

11.5 Robustness of the Toric Code Phase of Matter – Example of Topologically Ordered Matter

The excitation gap in of the toric code "protects" it from small perturbations and changes in the Hamiltonian. Indeed, the phase is "robust" against any small variations in the details of the Hamiltonian. To see this, let us suppose we have

$$H = H_{toric\,code} + \lambda \delta H$$

where H is the toric code Hamiltonian defined above, and δH is some arbitrary Hamiltonian (with local terms only) and λ is some small parameter. The claim is that for small enough λ, the topological properties of the phase of matter (such as the 4-fold degenerate ground state, and the exitations with their braiding statistics) will remain unchanged.

The easiest fact that we can test is that the four ground states remain robust and unmixed by the perturbation. To see this, let us pick some particular form for the δH such as a sum of σ^x on all edges

$$\delta H = \sum_i \sigma_i^x$$

(we will realize that the actual form we choose won't matter for the argument we make here). Now let us treat δH in perturbation theory. In the absence of the perturbation, we have four ground states $|\psi_{ee}\rangle$, $|\psi_{eo}\rangle$, $|\psi_{oe}\rangle$, $|\psi_{oo}\rangle$. Then if we add the perturbation order by order to one of these ground states, qualitatively we obtain[6]

$$|\tilde{\psi}\rangle = |\psi\rangle + (G\,\delta H)|\psi\rangle + (G\,\delta H)^2|\psi\rangle + \ldots$$

and the energy modified by the perturbing Hamiltonian is then

$$E = \langle\tilde{\psi}|H_{toric} + \delta H|\tilde{\psi}\rangle$$

where here G is the greens function, which includes an energy denominator at least as big as the excitation gap Δ, so that successive terms in the expansion are smaller by order λ/Δ. The point here is that at M^{th} order in perturbation theory, we can only generate wavefunctions that differ from the original ground state by M applications of δH. Now recall that one cannot even distinguish the ground state sectors from each other unless one has a string operator that wraps all the way around the torus. Thus, the result of this calculation is identical for the four ground states out to very high order of perturbation theory, and any splitting of the four ground state sectors (or any mixing of the sectors) will be suppressed exponentially as $(\lambda/\Delta)^L$ which can be made arbitrarily small for a big system. It is clear that this general argument is not specific to the particular form of δH we have chosen.

One can go further and ask what happens to the excited particles when a perturbation is applied to the system. Similarly, we can perform a perturbation series. Here what happens is that the

[6] This is a Brillouin-Wigner perturbation theory, where successive terms are rigorously λ/Δ smaller.

particles — which started as point defects — develop a nonzero length scale. As one moves a distance x further away from the particle, the influence of the presence of that particle decays as $(\lambda/\Delta)^x$. Again, if λ is small, then from a sufficiently far distance away, the particle again looks like a point. In particular, if one particle is braided around another at a sufficient distance away, it accumulates the expected phase that the pure toric code would have predicted. There are several strong arguments for this. First, we can explicitly write an expression for the braiding phase and show that the corrections do indeed drop exponentially by exactly the same arguments. Secondly, we recall the idea of rigidity presented in section 8.4 — it is not possible that the braiding phases in a theory change an arbitrarily small amount.

11.6 The Notion of Topological Order

The type of protection from small perturbations that we have just discovered is the basis for a very useful definition of topological order. A topologically ordered system will have multiple degenerate ground states when put on a surface with nonzero genus (i.e., a a torus, or a system with a hole cut in it) which we call $|\psi_\alpha\rangle$. To have topological order we should expect

$$\langle\psi_i|\text{any local operator}|\psi_j\rangle = C\delta_{ij}$$

where C depends on the particular operator and there may be corrections that are only exponentially small in the size of the system. In other words, the multiple ground states locally look just like each other, but are mutually orthogonal.

Chapter 12

Kitaev's Generalized Toric Code: The Quantum Double of a Group — Lattice Gauge Theory

Kitaev constructed an ingenious way to build a topological model from an arbitrary group G on a lattice. This is very much the generalization of the toric code, except that instead of using simple spins on edges, we give the edges values of elements of the group. The construction is based on lattice gauge theory, and will include the toric code as a simple example, where the group is $\mathbb{Z}_2$, the group with two elements[1].

We begin by defining a graph (which could be a regular lattice, or could be disordered). We define an orientation to each edge as an arrow as given in Fig. 12.1

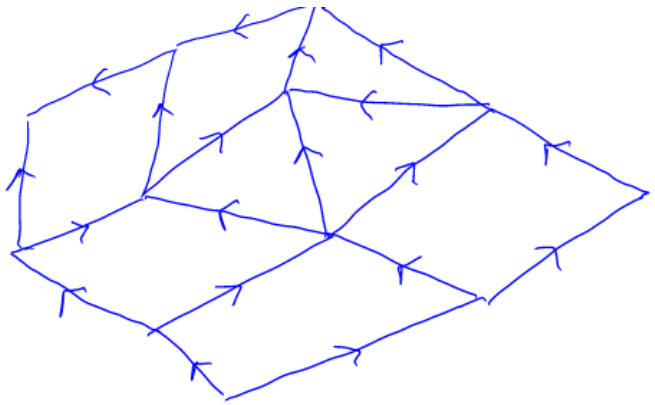


Figure 12.1: Part of a directed graph.

We choose a group G with group elements $g \in G$. The Hilbert space is defined by labeling edges with the group elements g. Inverting the arrow on an edge has the effect of inverting the group elemetn $g \to g^{-1}$ as shown in Fig. 12.2.

[1] I present this model on the "dual" graph compared to Kitaev's presentation.

Figure 12.2: Inverting the direction on an edge inverts the group element.

We now define a vertex operator V_α for a vertex α with all arrows pointed in as a projector which enforces that the product of group elements around the vertex to be the identity e, as shown in Fig. 12.3. This is the string-net vertex fusion rule.

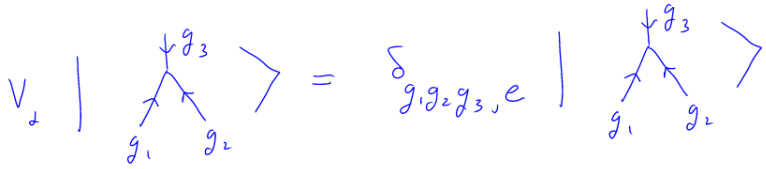


Figure 12.3: Definition of V_α when all arrows are directed into the vertex (if a vertex is directed out, one can invert the arrow and invert the group element). The vertex operator gives zero unless the product of group elements around the plaquette gives the identity element e

We can then define a plaquette operator $P_\beta(h)$ to premultiply the (clockwise orientied) group elements around a plaquette β by the group element h, as shown in Fig. 12.4.

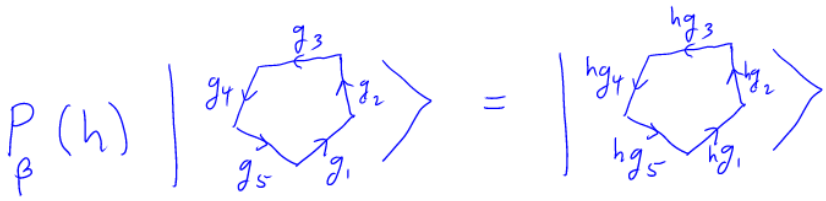


Figure 12.4: The plaquette operator $P_\beta(h)$ premultiplies all of the clockwise oriented bonds by the element h.

The total plaquette operator that (the one that will enter the Hamiltonian) is then defined to be

$$P_\beta = \sum_{g \in G} P_\beta(g)$$

It is easy to see that the plaquette operator and the vertex operator commute.

Relation to toric code

How does this related to the toric code? Consider the group $\mathbb{Z}_2$ of two elements where we write the two elements as $\{1, -1\}$. We can think of these as being spin up and spin down on the lattice. Since

$g = g^{-1}$ for every element we don't need to put arrows on the lattice.

$$P_\beta(1) = \text{identity operator}$$
$$P_\beta(-1) = \text{multiply all edges by -1. (i.e. flip all edges)}$$

and we have

$$P_\beta = P_\beta(1) + P_\beta(-1)$$

whereas the vertex operator is given by

$$V_\alpha = \begin{cases} 1 \text{ if an even number of edges are spin down} \\ 0 \text{ if an odd number} \end{cases}$$

we see that (up to the constants being added which are not interesting) these are simply the toric code vertex and plaquette operators.

The generalization of the toric code to theories built on the group $\mathbb{Z}_n$ (group of integers under addition modulo n) is rather straightforward, and also results in an abelian TQFT. The electric and magentic particles then have $\mathbb{Z}_n$ fusion rules instead of $\mathbb{Z}_2$ as in the toric code. We can think of this still as being a string net — with the new string net fusion rules at the vertex being now given by the structure of the group G.

The generalization to nonabelian groups is more nontrivial, and requires some amount of group theory to understand. The resulting TQFT is known as the quantum double (or Drinfeld double) of the group. The particles types of the TQFT are given by (C, χ) where C is a conjugacy class and χ is an irreducible representation of the centralizer of the conjugacy class[2]. Generically one will have nonabelian anyons. I will not go through this argument in detail. See Kitaev for more.

This model by Kitaev is essentially a lattice gauge theory. Essentially the wavefunction is given by a unique state plus everything that is "gauge equivalent" (meaning can be obtained by plaquette flips). Let us think in terms of the dual lattice for a moment (so plaquettes become dual-vertices and vertices become dual-plaquettes). The sum over group elements of $P_\beta(h)$ enforces gauge invariance of the theory at the dual vertices. The vertex operator V_α then assures there is no magnetic flux penetrating the dual plaquette.

[2]Two elements g and h of a group are called conjugate if $g = uhu^{-1}$ for some u in the group. A conjugacy class is a set of elements of a group that are all conjugate to each other. A group is naturally partitioned into nonintersecting conjugacy classes. A centralizer of an element g is the set of all elements of the group u that commute with it $ug = gu$.

Chapter 13

More Generalizing The Toric Code: Loop Gases and String Nets

The general ideas presented with the toric code can be further generalized topologically ordered phases of matter. They key generalizations were made by Levin and Wen. Also we will discuss in some of the language of the work of Freedman et al. And for the doubled fibonacci model, Fidkowski et al.

A key idea is that the underlying lattice is not very crucial to the details of the toric code. Indeed, we can write the toric code on any lattice structure and even on an irregular lattice, so it is often useful to dispense with the lattice altogether. This simplifies a lot of the thinking and allows us to generalize the model fairly simply. In fact it will allow us to manipulate our loop gas using the same sort of diagrammatic algebra we have been using all along! If we want to put the model back on a lattice at the end of the day, we can do this (we show an example in the double semion model) although it can start to look a bit more ugly.

13.1 Toric Code Loop Gas

We start by abstracting the toric code to simply a gas of fluctuating non-intersecting loops — no longer paying attention to a lattice. An example of a loop gas configuration is shown in Fig. 13.1

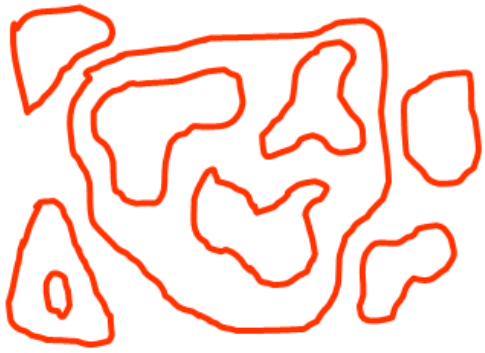


Figure 13.1: A loop gas in 2d. We can think of this as particle world-lines in 1+1 d.

Note, since this is in 2d, there are no over and under crossings — we can think about this picture as being some sort of world-lines for particles in 1+1d.

We can write the toric code wavefunction in the form of

$$|\psi\rangle = \sum_{\substack{\text{all loop configs that can be} \\ \text{obtained from a reference} \\ \text{loop config}}} |\text{loop config}\rangle \tag{13.1}$$

Where the types of "moves" one can make are similar to the diagrammatic moves we have been discussing for world lines in 2+1 d previously.

Move 1: "Isotopy" = smooth deformation of a loop. As shown in Fig. 13.2. We have always allowed smooth deformations in our diagrammmatic algebras.

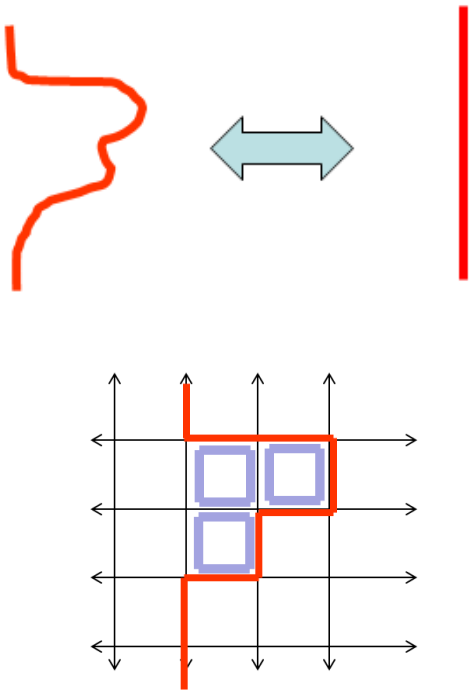


Figure 13.2: **Isotopy** (Top) Off the lattice this is just deformation of a line. (Bottom) on the lattice, this is implemented by flipping over the blue plaquettes.

Move 2: "Adding or removing a loop". As shown in Fig. 13.3

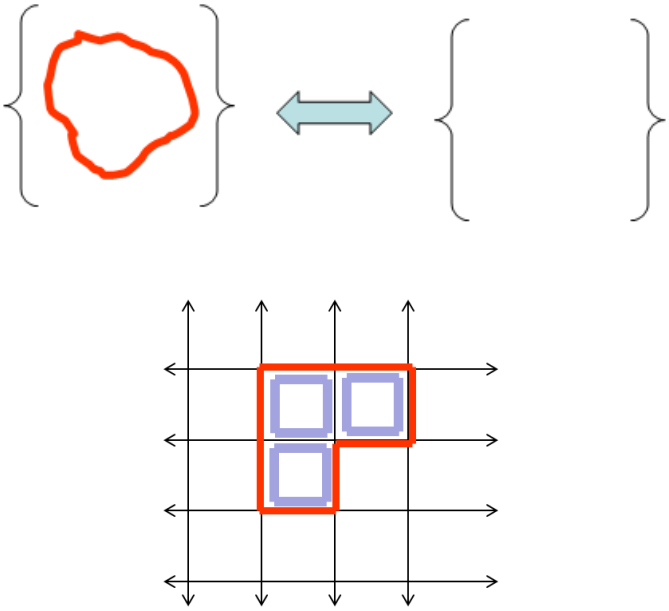


Figure 13.3: **Adding or Removing a loop** (Top) Off the lattice (Bottom) On the lattice we flip the shown plaquettes.

Move 3: "Surgery" or reconnection of loops. As shown in Fig. 13.4

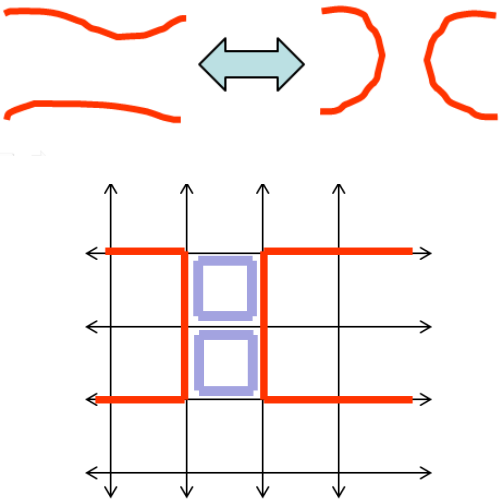


Figure 13.4: **Loop Surgery** (Top) Off lattice surgery (Bottom) On lattice, flip the shown plaquettes

We can summarize these rules with simple skein-like relations as shown in Fig. 13.5

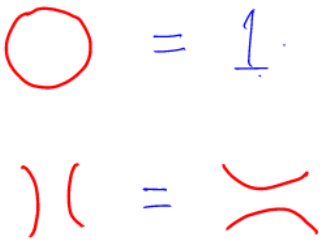


Figure 13.5: "Skein" relations for the toric-code loop gas. The unity on the right of the top line means that the amplitude in the superposition that forms the wavefunciton is unchanged (multiplied by unity) under removal or addition of a loop.

The ground state obviously decomposes into four sectors on a torus depending on the parity of the number of loops going around the handles of the torus.

13.1.1 Excitations of the Loop Gas

An end of a string in a loop gas corresponds to some sort of excitation (like a vertex excitation on the lattice). However, on the lattice, the vertex excitation could be either e or f, so how do we distinguish these off the lattice?

First we note that the string can end in many ways as shown in Fig. 13.6.

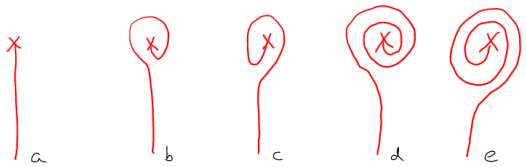


Figure 13.6: Ends of strings can be wrapped either way, and multiple times. a and b are different, c is equivalent to b by surgery. Similarly d and e are both the equivalent to a.

However, it turns out, due to the surgery rule, that there are actually only two inequivalent endings, a, and b from this list. To see this

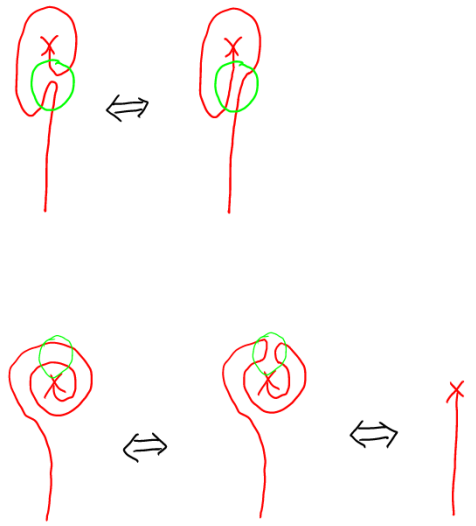


Figure 13.7: Loop equivalences. Surgery is done inside the light green circles. The final equality on the lower right is just pulling the string tight.

We now attempt to figure out the nature of these excitations by applying the twist operator $\hat{\theta}$ which rotates the excitation by 2π. This rotation wraps an untwisted particle's string into a loop as shown in Fig. 13.8

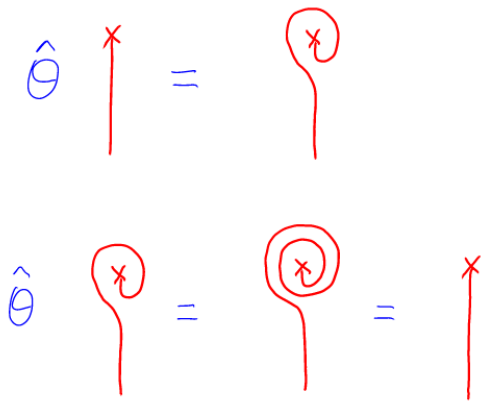


Figure 13.8: Rotation

From these relations we can determine that the eigenvalues of the rotation operator are $+1$, correpsonding to the e particle and -1 corresponing to the f particle, as shown in Fig. 13.9.

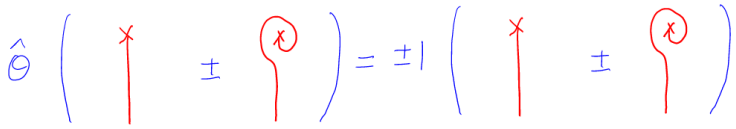


Figure 13.9: The eigenvectors of the rotation operator $\hat{\theta}$

Thus, the electric particle is the superposition of a straight line and a twisted line. This may seem surprising, because on the lattice it seems that we can make a pair of e particles flipping a single bond, which might seem like just a straight line between the two endpoints. However, we must also consider the possibility that the endpoint is surrounded by a loop when the defect line is created!

The magnetic particle m can be constructed by fusing together $e \times f$. The result should be the same as our prior definition of the magnetic particle. Recall that the ground state should be a superposition of no-loop and loop (with a positive sign). This is what we learned from considering a plaquette operator to be a minimal loop. If we take a superposition with a minus sign, we get something orthogonal to the ground state, which should be the magnetic particle, as shown in Fig. 13.10.

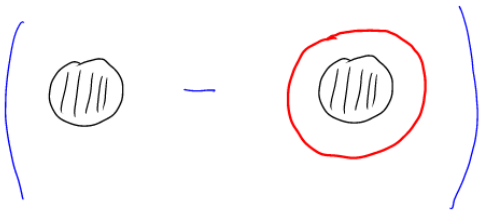


Figure 13.10: The black disk is some region of our model. Forming a superposition of this region, and this region with a loop around it, with a minus sign between the two pieces, must be orthogonal to the ground state — it puts a magnetic excitation m in the region.

13.2 The Double Semion Loop Gas

A rather minor modification of the skein rules for the loop gas results in a somewhat different topological phase of matter. Consider changing the rules so that each loop removal/addition, and each surgery, incurs a minus sign. Note that these two minus signs are consistent with each other because each surgery changes the parity of the number of loops in the system.

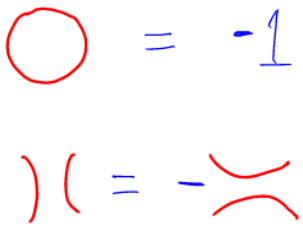


Figure 13.11: "Skein" relations for the double-semion loop gas. Each loop removal/addition and each surgery incurs a minus sign. Note that these are the same as the Kauffman rules when we considered semions.

Note that these rules were precisely the skein rules we used for the Kauffman invariant when we considered semions!

From these rules we expect wavefunctions of the form

$$|\psi\rangle = \sum_{\substack{\text{all loop configs that can be} \\ \text{obtained from a reference} \\ \text{loop config}}} (-1)^{\text{Number of Loops}} |\text{loop config}\rangle \tag{13.2}$$

We can think of the prefactor (-1) to the number of loops, as being the wavefunction written in the basis of loop configurations.

As with the toric code, there should be four ground states on the torus corresponding to the different possible parities around the two handles.

13.2.1 Microscopic Model of Doubled Semions

We now turn to try to build a microscopic hamiltonian for the doubled semion loop gas. First, however, we realize that there is a problem with constructing this on a square lattice. When four red lines touch at a corner we cannot tell if we have a single loop or two loops (See right of Fig.13.12). To avoid this problem we switch to using a trivalent network (the word "lattice" is not really appropriate, despite the fact that most people in condensd matter would call it a trivalent lattice). The simplest trivalent network is the honeycomb.

Honeycomb's Good

A rather trivial generalization is to change the lattice to a honeycomb as shown in Fig. 13.12. The advantage of this structure is that loops cannot intersect as they can (at the 4-fold corner) on the the square lattice.

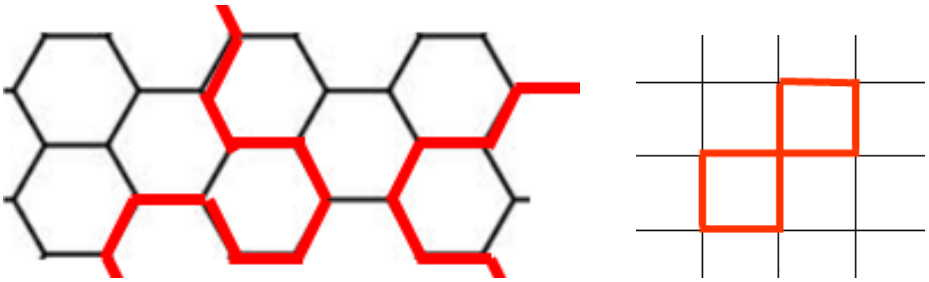


Figure 13.12: Left: Toric code on a honeycomb, loops are nonintersecting. Right: On the square lattice loops can intersect at corners and one cannot tell if this picture represents one loop or two.

As in the previous square case, the vertex operator must assure that an even number of red bonds intersect at each vertex, and the plaquette operator now flips all six spins around a plaquette.

In fact, any trivalent network will be suitable. In all cases the vertex operator enforces that we are considering only loop gases – now with no self-intersections allowed. The plaquette operators will flip all of the bonds around a plaquette, as in the toric code, but will now assign signs such that creating or destroying a loop incurs a minus sign.

To see how this can be achieved consider Fig. 13.13. Depending on the initial state, when the plaquette is flipped, one may or may not obtain a minus sign.

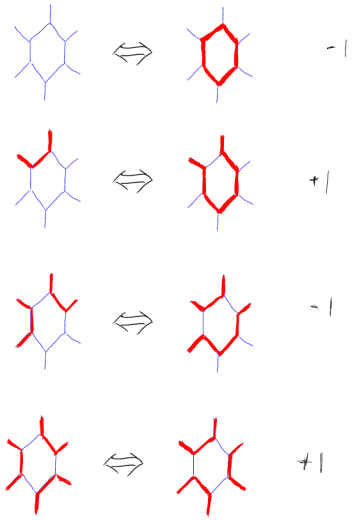


Figure 13.13: Some plaquette flips for the double semion model on the hexagon. The top line obviously adds a loop, so should get a minus sign. The second line just stretches a loop over a plaquette, so does not get a minus sign. The third line is a surgery so gets a minus sign. The fourth line is a double surgery, so gets no minus sign.

One way of determining if one should or should not get a minus sign is to count the number of red bonds touching the outside of the hexagon (sometimes called the outside "legs"). Because red bonds form closed loops, the number of red legs of a hexagon must be even. If the number of red legs is a multiple of four, then one gets a minus sign in the flip.

One can thus write a plaquette operator for the hexagon as

$$P'_\beta = \left(\prod_{i \in \, plaquette \, \beta} \sigma_i^x \right) \; (-1)^{\frac{1}{4} \sum_{j \in \, legs \, of \, \beta} (\sigma_j^z + 1)}$$

The overall Hamiltonian for this model is then

$$H = -\sum_{vertices\ \alpha} V_\alpha - \sum_{plaquettes\ \beta} P'_\beta$$

This Hamiltonian was first written down by Levin and Wen.

13.2.2 Double Semion Excitations

The addition of the sign in the surgery rule changes the effect of rotations. We now have the added sign in Fig. 13.14

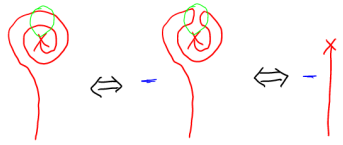


Figure 13.14: Surgery incurs a minus sign. Compare to fig. 13.7

Resulting in the effect of rotation being Fig. 13.15

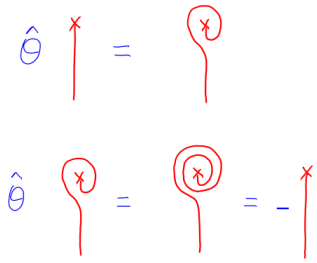


Figure 13.15: Surgery incurs a minus sign. Compare to fig. 13.7

Again we can use these to give us the eigenstates of the rotation operator as shown in Fig. 13.16

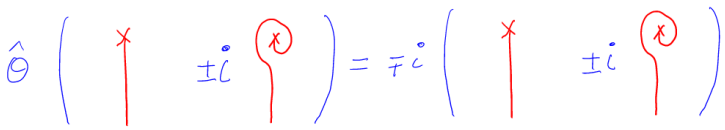


Figure 13.16: Eigenstates of the rotation operator for the doubles semion model.

Thus we have two particle types with twist factors i and $-i$. These are right and left-handed semions. It is interesting that we used the skein rules for a model of semions to build our loop gas, and we got out two types of particles — **Both** right and left handed semions. This is perhaps to be expected, since nowhere in our input rules did we ever break "time-reversal" or say whether the original theory was right or left handed — it comes out to be both!

As with the toric code, there is also a magnetic particle which can be thought of as a fusion between the left and right handed particle — or could just be considered as a superposition analagous

to Fig. 13.10, except now with a plus sign (since the ground state now is a superposition with a minus sign, being that a loop addition now incurs a minus sign). Thus the duouble semion model has four particles I, ϕ, ϕ^*, m where ϕ and ϕ^* are the right and lefthanded semions. The full fusion rules are

$\times$	I	ϕ	ϕ^*	m
I	I	ϕ	ϕ^*	m
ϕ	ϕ	I	m	ϕ^*
ϕ^*	ϕ^*	m	I	ϕ
m	m	ϕ^*	ϕ	I

Quantum Doubling: We emphasize again that we started with a theory having the kauffman rules of a model of semions (but we did not need to put in the braiding by hand) and we got out a theory that has both right and left handed semions. This priniciple is very general. If we start with any theory of anyons and build a quantum loop gas from it (not putting in any of the braiding relations) we will get out the *doubled* theory, meaning it has both right and left handed versions of the theory.

As mentioned above the ground state should be thought of a the positive eigenstate of the operator shown in Fig. 13.10 (including the minus sign). Note that this combination of identity minus the string with a prefactor of $1/\mathcal{D} = 1/\sqrt{2}$ is precisely the Ω strand (or Kirby color) of the original semion theory (which has only two particles, the identity or vacuum, and the semion or single string)[1] If we think in three dimensions, the ground state is defined as having no flux through any loops.

13.3 General String Net

Given our success with the loop gases, we would like to generalize the idea to more general so-called "string-nets". In the case of the double semion model as discussed above, we can really think of the loops as being particle world-lines living in the plane (but with no crossings allowed). We would like to upgrade this idea to a set of world-lines, still living in a plane, but where different types of particles are allowed, and they can fuse and split (but again, we allow no braiding). This type of multi-valued loop gas should look familiar from Kitaev's generalized toric code, although the construction here is more general still since the edge labels need not form a group.

Thus in these string net models, we allow branching of loops, and we allow strings of different colors as shown in Fig. 13.17. We can think of this as being similar to the fusion diagrams we have encountered before – the allowed branchings being given by the allowed fusions of the string types. (We do not allow strings to go over or under each other though!).

Figure 13.17: A general string net, that allows branching, here with two colors.

[1] To check that this is indeed the Kirby color, show that a loop of this Kirby string will annihilate a flux going through the loop as in Section 10.4, and gives $\mathcal{D}$ on the vacuum.

We would like to similarly define a wavefunction to be of the form

$$|\psi\rangle = \sum_{\substack{\text{string} \\ \text{nets}}} \Phi(\text{net config}) \ \ |\text{net config}\rangle$$

where the prefactors Φ(net config) satisfy some graphical rules as shown in Fig. 13.18.

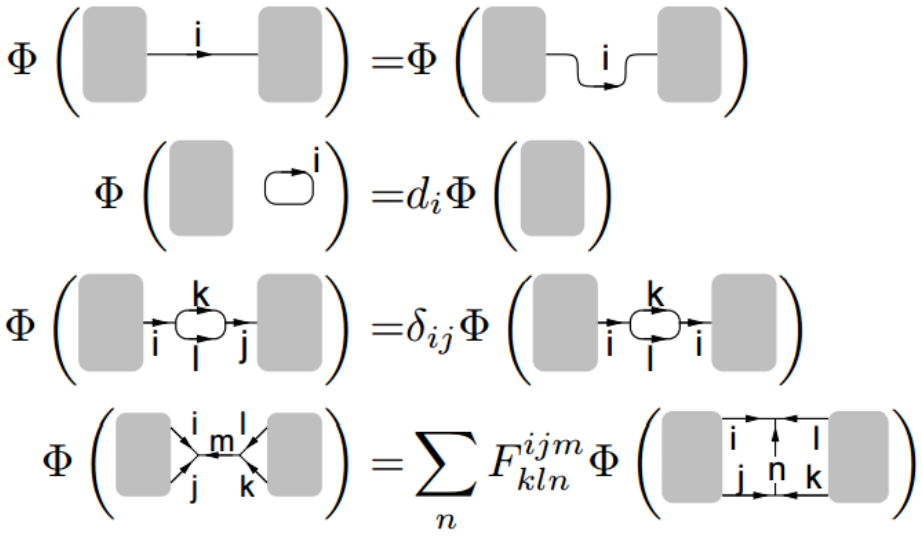


Figure 13.18: Rules for a string net. The grey regions are meant to be the same on both the left and the right of the diagram. Figure stolen from Levin and Wen.

The meaning of these rules are as follows: The first rule is simply saying that we can deform one of the strings without chaning the value of the prefactor Φ. The second rule says that removal of a loop multiplies the prefactor Φ by a constant which we call the quantum dimension of the loop d_a. The third rule is just our "locality" principle — if a quantum number i enters a region, that quantum number must also come out of the region. This rule is irrelevant in the case of teh the toric code and the double semion theory, because loops are not allowed to branch. The final rule is a more complicated one which allows for the possibility of making an "F-move" on a diagram – relating the prefector on the left to a sum of prefactors of diagrams on the right. The analogue F move in the toric code and double semion model are the second lines of Fig. 13.5 and Fig. 13.11.

It is important to note that the F-matrix used to define define the string net (last line of Fig. 13.18) must satisfy the pentagon equations for consistency. It is crucial to note that one need not have define any R matrices, since the string net model is defined entirely in 2d without having any crossings of strings — so the F matrices do not have to correspond to an actual anyon theory. The theory that results is known as a Drinfeld double or quantum double.

Note however, certain F-matrices do have corresponding R matrices which solve the hexagon equations. In this case, it is possible to think of the string net model as being built from an underlying anyon theory — the resuling topological theory is the simple "double" of the underlying anyon theory (i.e, just a right handed and a left handed copy of the theory). The ground state will then be the $\mathcal{D}$ eigenstate of the Kirby color loop – which makes it fairly easy to write a Hamiltonian on a lattice for this string net model.

13.4 Doubled Fibonacci Model

As an example, let us try to build a string net model from from the Fibonacci anyon theory. Again we will not put in the braiding information, we only put in the fusion algebra.

We will write the identity (or vacuum) particle as no-line and the fibonacci particle τ as a red line, Since $\tau \times \tau$ can fuse to τ we expect that this loop gas will allow our (red) loops to branch. We thus call this version of a loop gas a "string net" (or a branching loop gas) as in Fig. 13.19.

Figure 13.19: A branching string net for the doubled Fibonacci model.

Starting with Eq. 7.2, we consider the following F-moves as shown in Fig. 13.20

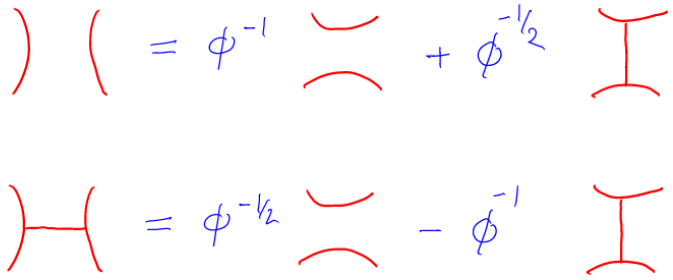


Figure 13.20: Rules for building the doubled fibonacci model.

Where here $\phi = (1 + \sqrt{5})/2$ and (the values of these coefficients come from the values of the F-matrix in Eq. 7.2.

We also expect to have rules of the form of Fig. 13.21

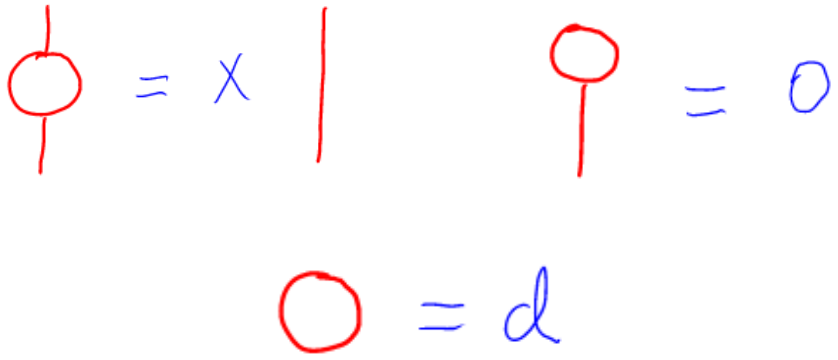


Figure 13.21: Rules for building the doubled fibonacci model.

The first and second rules[2] are results of locality. The final rule is the usual rule that a loop can be removed and replaced by a number. This final rule also tells us that the ground state should be a $\mathcal{D}$ eigenstate of the Kirby string operator — since the Kirby Ω string is a sum of $1/\mathcal{D}$ times the identity operator and $d/\mathcal{D}$ times a loop of τ, whose value is now d, adding a Kirby string give $1/\mathcal{D} + d^2/\mathcal{D} = \mathcal{D}$

We can then pin down the values of d and X in these equations. To do this, we connect the strings on the right of Fig. 13.20 to give Fig. 13.22.

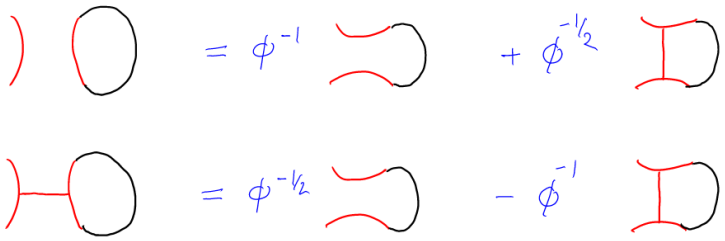


Figure 13.22: Starting with Fig. 13.20 and closing strings to the right hand. The black strings should be imagined to be red — they are drawn black so one can see what is added compared to Fig.!13.20

Using the laws above we these equations are translated to

$$d = \phi^{-1} + \phi^{-1/2} X$$
$$0 = \phi^{-1/2} - \phi^{-1} X$$

which we solve to obtain

$$X = \phi^{1/2}$$
$$d = \phi^{-1} + 1 = \phi$$

The fact that $d = \phi$ is not surprising being that this is the expected quantum dimension for a Fibonacci particle.

[2] In fact we can prove that the tadpole rule must be zero. This is a homework problem!

With the values we obtain for X and d, we now have a full set of rules in Fig. 13.20 and 13.21. We can then write a ground state wavefunction of the form

$$|\psi\rangle = \sum_{\substack{\text{all string net configs that} \\ \text{can be obtained from a ref-} \\ \text{erence config}}} \Phi(\text{net config}) \ |\text{net config}\rangle$$

This looks quite similar to our above toric code loop gas, except now we allow branching string nets instead of just loops, and also the kets have a prefactor ϕ. These prefactors are chosen such that the algebraic rules described above are satisfied. I.e., removing a loop increases Φ by a factor of d. Removing a bubble (as in the upper left of 13.21) increases Φ by a factor of X. Then F tell us the relationship between three values of Φ where changes in the diagram are made as shown in Fig. 13.20.

13.4.1 Excitations

As with the double-semion model we should be able to determine the quasiparticle eigenstates by looking at how a single line can end in a defect. We claim that all possible line endings can be reduced, by F-moves, to one of the three possible endings shown in Fig. 13.23

Figure 13.23: Possible string endings in the doubled fibonacci string net model.

Just as an example, consider the ending shown on the left of Fig. 13.24. By using an F-move, it is reduced to a combination of the three presented above.

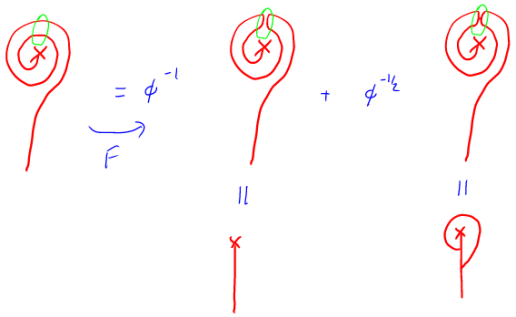


Figure 13.24: An example of reducing a more complicated string ending into one fo the three endings shown in Fig. 13.23.

As in the case of the toric code and the double semion model, we can figure out the twist factors by rotating these diagrams as shown in Fig. 13.25 and then using F-matrices to reduce the result back to linear combinations of the same three possible endings.

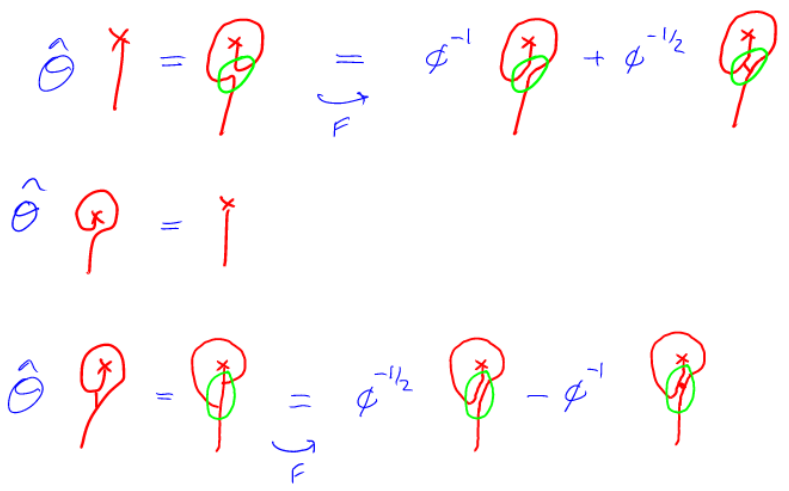


Figure 13.25: The rotation operator $\hat{\Theta}$ applied to the possible string endings. Then using F matrices we reduce the results to linear combinations of the same endings.

We can write these diagrammatic equations more algebraically by

$$\hat{\Theta}\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 & \phi^{-1} & \phi^{-1/2} \\ 1 & 0 & 0 \\ 0 & \phi^{-1/2} & -\phi^{-1} \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

The eigenvectors of this matrix are the particle types with definite twist factrors given by their eigenvalues under rotation.

With a bit of algebra it can be shown that the eigenvalues of this matrix are given by

$$\theta = e^{i\pi 4/5}, \quad e^{-i\pi 4/5}, \quad 1,$$

The first two correspond to the expected spin factors for a right-handed fibonacci anyon τ or left-handed fibonacci anyon τ^* (recall that we worked out the spin factor using the hexagon equation earlier. See 8.3.). The final possibility represents the fusion of these two objects $\tau \times \tau^*$. Indeed, these are all of the possible particle types in the doubled-fibonacci theory. Since the theory was based on a full anyon theory with braiding fully defined, we expected to get both a right- and left-handed copy of the Fibonacci model and indeed we did. (We never broke time reversal in the definition of the model so we should get both hands of the theory!).

13.4.2 Ground State Degeneracy

It is a bit tricky to figure out the ground state degeneracy here. Using the above skein rules, any configuration can be reduced to a linear combination of four simple configuation – corresponding to the possibilities of having a loop, or not having a loop, around each handle. An example of reducing two loops around a handle to a linear combination of zero and one loop is given in Fig. 13.26

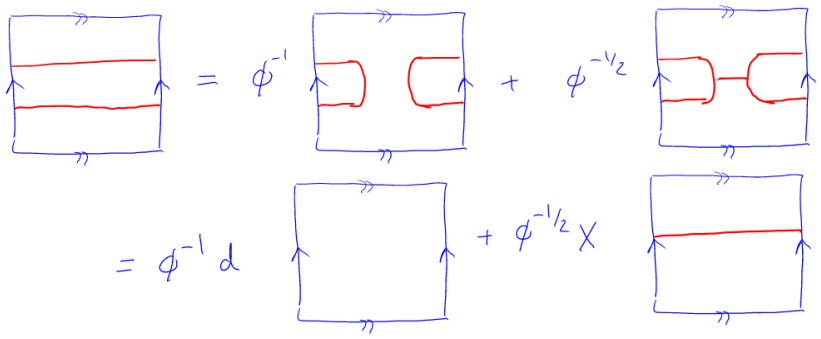


Figure 13.26: Reducing two loops around a handle to a linear combination of one loops and zero loops.

13.5 Add details of Levin Wen Model on the lattice?

13.6 Appendix: S-matrix for Fibonacci Anyons

Without doing much work, we can figure out the S-matrix for Fibonacci anyons. There are only 2 particles in the theory I and τ. Further we know that the quantum dimension of τ is $\phi = (1+\sqrt{5})/2$. Thus, the total quantum dimension is $\mathcal{D}^2 = 1 + \phi^2 = 2 + \phi$ and the S matrix must be of the form

$$S = \frac{1}{\mathcal{D}} \begin{pmatrix} 1 & \phi \\ \phi & y \end{pmatrix}$$

where the constraint of unitarity immediately fixes $y = -1$.

We can check this by using F and R matrices to determine the value of two linked rings explicitly as shown in Fig. 13.27

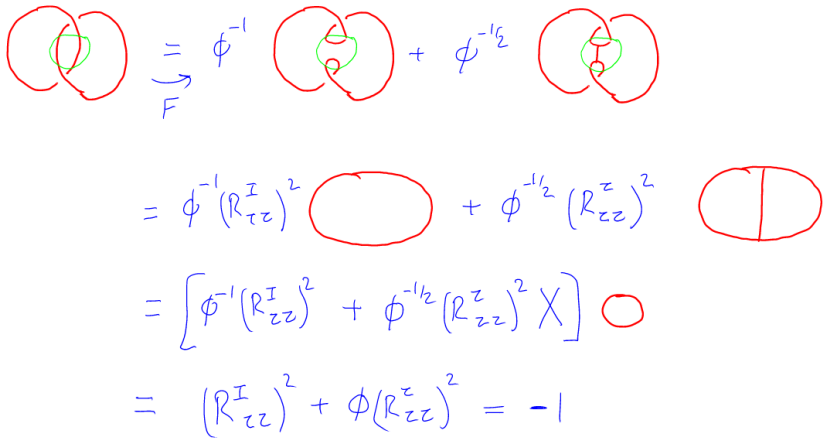


Figure 13.27: Calculating the nontrivial element of the Fibonacci anyon S-matrix.

Chapter 14

Introduction to Quantum Hall — The Integer Effect

The fractional quantum Hall effect is the best studied of all topologically ordered states of matter. In fact it is the *only* system which is extremely convincingly observed to be topologically ordered in experiment[1]. We will thus spend quite a bit of time discussing quantum Hall effects in detail. Before we can discuss fractional quantum Hall effect we need to discuss the basics, i.e., the integer quantum Hall effect.[2]

14.1 Classical Hall Effect

In 1879 Edwin Hall discovered that when a current is run perpendicular to a magenetic field, a voltage is generated perpendicular to both field and current, and proportional to both (See Fig. 14.1). This voltage is now known as the Hall voltage. Drude theory, treating a metal as a gas of electrons, explains the Hall voltage as being a simple result of the Lorentz force on electrons.

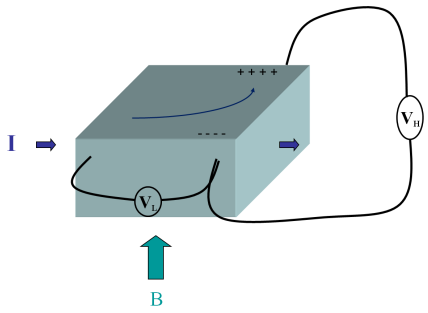


Figure 14.1: Hall voltage V_H perpendicular to both magnetic field and current, and proportional to both. Also one measures a longitudinal voltage in the same direction as the current, roughly independent of magnetic field.

[1]There are a good number of other contenders now. Probably the most convincing other case is ^{3}HeA phase 2d films. Although very few experiments have actually been done on this. Other strong contenders include Majorana wires, certain exotic superconductors, and a few frustrated quantum spin systems.

[2]There is quite a bit further to be learned from integer quantum Hall effect as this is the simplest case of a so-called topological insulator. While there are obviously strong overlaps between the ideas of topological insulators and our explorations of topological quantum field theories, this connection would take us too far afield, so we defer this for another time.

14.2 Two-Dimensional Electrons

In the late 1960s and early 70s semiconductor technology made it possible to do experiments with electrons that live in two dimensions. First MOSFETs[3] and later quantum wells were used to provide a confining potential for electrons in one direction, leaving motion only in the two remaining dimensions. As an example we will consider a quantum well structure, which is layered in the $\hat{z}$ direction as shown in Fig. 14.2.

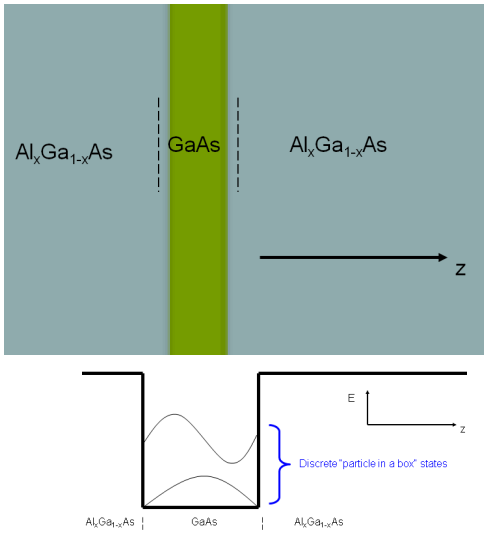


Figure 14.2: **Top** A quantum well structure is a quasi-two-dimensional layer of one semiconductor sandwiched between two other semiconductors. **Bottom** The potential felt by an electron is like a particle in a box. If the energy is low enough, the electron is stuck in the lowest particle-in-box wavefunction $\varphi_0(z)$ giving a total wavefunction $\Psi = \varphi_0(z)\psi(x, y)$ and having strictly two dimensional motion.

The electron moving in the z-direction experiences a strong confinement, such as the particle-in-box confinement shown in Fig. 14.2. The wavefunction of the electron then takes the form $\varphi(z)$ in the z-direction. If the energy (i.e. the temperature and coulomb interaction) is very low compared to the gap between the particle-in-box states, then the electron is frozen in the lowest particle-in-box state $\varphi_0(z)$ and the total wavefunction of the electron is $\Psi(x, y, z) = \varphi_0(z)\psi(x, y)$ leaving only the x and y degrees of freedom. Thus we have a strictly two dimensional electron.

More recently two dimensional electronic systems have also been observed in single-layer atomic systems such as graphene. (Although even then, the same argument needs to be used — that the motion of the electron is "frozen" in the z-direction and only has freedom to move in x and y).

14.3 Phenomenology of Integer Quantum Hall Effect

In 1980 Klaus von Klitzing, having just left a postdoctoral position at Oxford, went to a new job at Grenoble carrying some new high mobility[4] two dimensional electron samples grown by (now Sir) Michael Pepper at Cambridge. He put them in high magnetic field and cooled them down to a few degrees Kelvin temperature where he discovered something very different from what Hall had seen a hundred years earlier. An example of this type of experiment is shown in Fig. 14.3.

[3] Metal Oxide Semiconductor Field Effect Transistors
[4] Meaning very clean

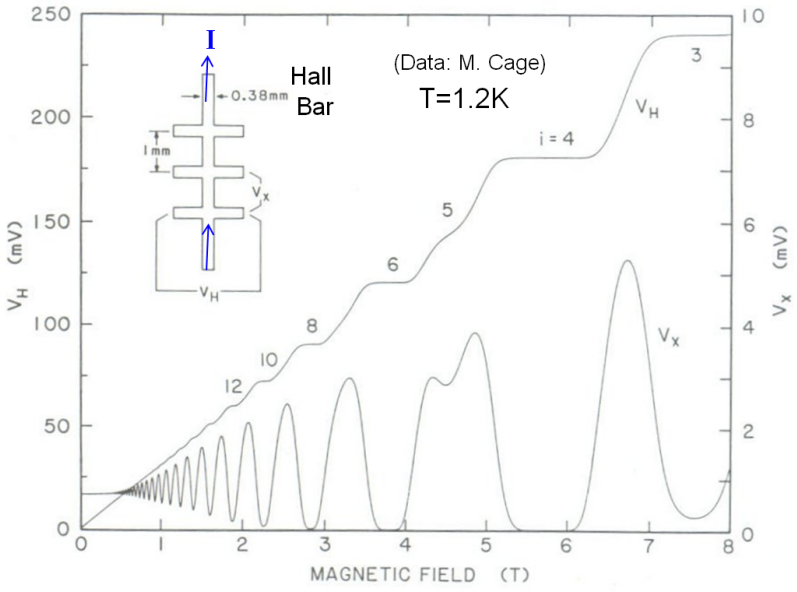


Figure 14.3: An example of an Integer Quantum Hall experiment. The plateaus in V_H are such that $V_H = (1/i)(h/e^2)I$ with i the integer displayed over the plateau — where h is Planck's constant and e is the electron charge. At the same magnetic field where a plateau occurs in V_H the longitudinal voltage drops to zero. Note that at very low field, the Hall voltage is linear in B and the longitudinal voltage is independent of B, as would be predicted by Drude theory.

At low magnetic field, the longitudinal voltage is relatively constant whereas the Hall voltage is linear in magnetic field — both of these are precisely what would be predicted by Drude theory. However, at high magnetic field, plateaus form in the Hall voltage with concomitant zeros of the longitudinal voltages. The plateaus have precisely the value

$$V_H = \frac{1}{i}\frac{h}{e^2}I$$

where I is the current, h is Planck's constant and e is the electron charge. Here i is an integer as shown in the figure. Or equivalently we have

$$R_H = \frac{1}{i}\frac{h}{e^2} = 1/G_H \tag{14.1}$$

with R_H the Hall resistance where G_H the Hall conductance. Where we have plateaus in the Hall voltage, we have zeros in the longitudinal voltage and resitstance

$$R_L = 0$$

which implies we have a dissipationless state — similar to a superfluid. These statements become increasingly precise as the temperature is lowered.

We should remember that conductivity and resistivities are both 2 by 2 matrices and are inverses of each other[5]. In this quantum Hall state, these matrices are both purely off-diagonal. Thus we have the interesting situation that both the diagonal part of the conductivity (the longtidinal conductivity) is zero, *and* the diagonal part of the resistivity (the longitudinal resistivity) is also zero.

The plateau $R_H = (1/i)(h/e^2)$ occurs near the magnetic field such that the so-called filling fraction ratio

$$\nu = \frac{n\phi_0}{B}$$

[5]These are 2 by 2 matrices because they relate the vector electric field **E** to the vector current **j**

is roughly the integer i. Here n is the 2d electron density and ϕ_0 is the quantum of magnetic flux

$$\phi_0 = h/e$$

When von Klitzing discovered this effect he noticed mainly that the plateaus in the Hall resistance are extremely precisely given by Eq. 14.1 and the plateaus are extremely flat. He submitted his manuscript to PRL claiming that this would be a useful way to make a new resistance standard [6,7]. In fact the result has been shown to be precise and reproducible to better than a part in 10^{10}. This is like measuring the distance from London to Los Angeles to within a fraction of a millimeter. This accuracy should be extremely surprising. The samples are dirty, the electrical contacts are soldered on with big blobs of metal, and the shape of the sample is not very precisely defined.

14.4 Transport in Zero Disorder

In strictly zero disorder it is easy to show that the longitudinal resistance is zero and the Hall resistance is precisely linear in the magnetic field. This is a simple result of Galilean/Lorentz invariance. Suppose we have a two dimensional disorder-free system of electrons in the x, y plane and a magnetic field $\mathbf{B} = B\hat{\mathbf{z}}$ in the $\hat{z}$-direction perpendicular to the plane. The Lorentz force on an electron will be

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

If we then boost into a moving frame where

$$\mathbf{v} = \frac{\mathbf{E} \times \hat{z}}{|B|}$$

in this new frame we obtain $\mathbf{F} = \mathbf{0}$, so the ground state must be stationary in this frame.

Then we boost back into the lab frame, and we obtain a current

$$\mathbf{j} = -en\mathbf{v} = \frac{-en\mathbf{E} \times \hat{z}}{|B|}$$

thus giving us

$$R_L = 0$$
$$R_H = \frac{B}{ne}$$

which is exactly the prediction that Drude would have made for a disorder free system.

While this calculation is rigorous even with the effects of quantum mechanics and interactions, it relies on having strictly zero disorder.

14.5 The Landau Problem

In order to understand quantum Hall effect, we should start by understanding the physics of a charge particle in a Magnetic field — a problem first studied by Landau. For simplicity we assume our electrons are spinless (indeed, the spins tend to be polarized by the magnetic field anyway.) We will consider an electron in the x, y plane, with a magnetic field of magnitude B in the z direction. We will assume the system is periodic in the y direction with length L_y, but opern in the x direction, with length L_y (i.e., we are working on a cylinder actually). We will eventually consider a small

[6] The referee (who we now know was Steve Girvin) mentioned that at the time they already had resistance standards which were better than his initial measurement of one part in 10^6, but proposed would be a uniquely good measurement of the ratio h/e^2. The paper was resubmitted proposing to use the effect as a precise measurement of the fine structure constant. The paper was accepted and the Nobel prize for von Klitzing followed in 1985.

[7] The quantum Hall effect is used as a metrological resistance standard, and it is proposed that the Ohm will soon be *defined* in terms of the result of quantum Hall experiments.

amount of disorder (as we showed above this is crucial!), but for now let us assume the system has no disorder.

The Hamiltonian is

$$H_0 = \frac{(\mathbf{p} + e\mathbf{A})^2}{2m}$$

where e and m are the electron charge and mass, and $\mathbf{A}$ is the vector potential. We then have to choose a particular gauge to work in. Later on we will want to work in symmetric gauge (there is a homework problem on this!) For now we will work in the so-called "Landau" gauge

$$\mathbf{A} = Bx\hat{y}$$

which does indeed satisfy

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

as desired. The Hamiltonian is thus

$$H_0 = \frac{1}{2m}\left((p_x^2 + (p_y + eBx)^2\right)$$

where $p_j = -i\hbar\partial_j$.

The Hamiltonian is then translationally invarient in the $\hat{y}$ direction, so we can write the wavefunction as

$$\psi(x, y) = \phi_{k_y}(x)e^{ik_y y}$$

and due to the periodicity in the y-direction, we have

$$k_y = \frac{2\pi n}{L_y}$$

for some integer n. Plugging in this form gives a familiar Schroedinger equation

$$\left(\frac{p_x^2}{2m} + \frac{1}{2}m\omega_c^2(k_y\ell^2 + x)^2\right)\phi_{k_y}(x) = E\phi_{k_y}(x) \tag{14.2}$$

where ℓ is the so-called magentic length

$$\ell = \sqrt{\hbar/(eB)}$$

and ω_c is the cyclotron frequency

$$\omega_c = eB/m.$$

We recognize this Schroedinger equation as being just a harmonic oscillator where the center of the harmonic potential is shifted to $x = -k_y\ell^2$. Thus the eigenenergies are of the usual harmonic oscillator form

$$E_p = \hbar\omega_c\left(p + \frac{1}{2}\right) \tag{14.3}$$

where p is an integer. These quantized energy states are known as Landau levels. The form of the wavefunction will be harmonic oscillator on the x direction and plane-wave in the y-direction as shown in Fig. 14.4.

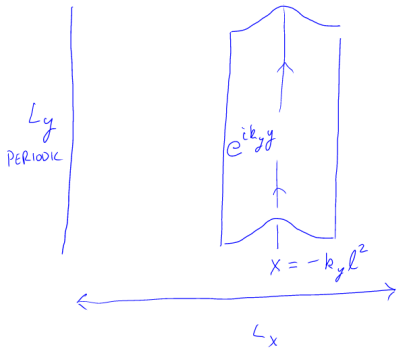


Figure 14.4: The shape of the wavefunction of an electron in a magnetic field using Landau gauge. The form of the wavefunction will be harmonic oscillator on the x direction and plane-wave in the y-direction

Fixing the energy by fixing p in Eq. 14.3, the value of k_y is quantized in units of $2\pi/L_y$. Further, the position x ranges over L_x, meaning that k_y ranges over L_x/ℓ^2. Thus the total number of possible values of k_y is

$$\text{Number of states in a Landau level} = \frac{L_x L_y}{2\pi\ell^2} = \frac{\text{Area} \quad B}{\phi_0}$$

where

$$\phi_0 = h/e$$

is the magnetic flux quantum. Thus, the number of states in a Landau level is equal to the number of magnetic flux quanta of magnetic field incident on the plane.

We can plot the density of states for electrons in a magnetic field, as shown in Fig. 14.5

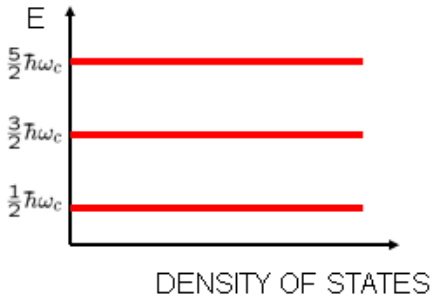


Figure 14.5: The density of states for spin-polarized (or spinless) electrons in a magnetic field. At energies equal to half-odd integer multiples of the cyclotron frequency, there is a spike of degenerate states, with degeneracy $\frac{\text{Area} \quad B}{\phi_0}$.

When there are multiple electrons present, we define the **filling fraction** to be the number of these Landau levels which are completely filled with electrons.

$$\nu = \frac{n\phi_0}{B}$$

where n is the density of electrons. Or equivalently we can write a relationship between the number of electrons in the system, N_e and the number of magnetic flux N_ϕ

$$N_e = \nu N_\phi$$

Incompressility of Integer Number of Filled Landau Levels:

When some integer number of Landau levels is filled, the chemical potential lies in the middle of the gap between the filled and unfilled states — analogous to a band insulator. In this case the the system is *incompressible*. This means there is a finite energy gap to creating any excitations — i.e., all excitations must involve removing an electron from a filled Landau level, promoting it above the energy gap to place it in an empty state. In particular excitations which change the density (compressions) are gapped. Further, at this precise integer filling fraction, the longitudinal conductivity is zero, and the Hall conductivity is precisely the quantized value $R_H = ne/B = (1/i)(h/e^2)$.

If we were to control the chemical potential in the experiment, we would have our answer as to why the Hall conductivity shows plateaus — for any value of the chemical potential, except for the special values $\mu = (\hbar\omega_c)(p + 1/2)$ with integer p, the electron number is pinned to $N = N_\phi/i$ where i is an integer, precisely i Landau levels are filled, there is a gap to excitations, and the Hall conductivity would be precisely quantized. However, in real experiments, it is actually the density that is fixed — which means that generically the chemical potential *does* sit in the degenerate band $\mu = (\hbar\omega_c)(p + 1/2)$ for some integer p and generically the filling fraction is tuned continuously and is not quantized.

Thus the incompressible state is very fine tuned. It occurs only for a very precise (integer) value of the filling fraction —for all other values of the filling fraction, some Landau level is partially filled and (at least neglecting interactions) the system would be extremely compressible, as there are many zero energy excitations corresponding to rearrangements of the electrons (which orbitals are filled and which are empty) within the partially filled Landau level.

While the system does have a gap under fine tuning, we will need something that will preserve the special properties of the fine tuned state even when we move away from the filling fraction which is precisely an integer. What does this is actually disorder — it will provide a reservoir for excess electrons (or holes) added (or subtracted) from the integer filled state. With disorder, the special properties of the quantized state are made robust.

What Does Disorder Do?

As mentioned above, we will need to add disorder to the system in order to achieved quantized Hall effect. What is the effect of this disorder? Disorder will spread out the energies in the band by having some regions where the potential is higher than average and some regions where the potential is lower than average. This spreads the sharp peak in the density of states into a broader band, as shown in Fig. 14.6.

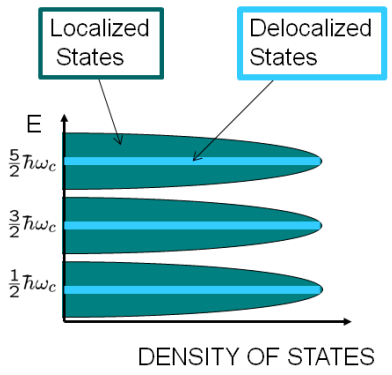


Figure 14.6: The density of states for spin-polarized (or spinless) electrons in a magnetic field with disoder. The Landau bands are spread out, with localized eigenstates in the tails and extended eigenstates near the middle.

Since current tends to flow perpendicular to potential gradients (i.e., it is hall current), eigenstates tend to follow contours of constant potential. Thus many of the eigenstates at high and low energy will be trapped in local minima or maxima — isolated in a hill or valley and circling the peak or bottom. The result is that the eigenstates in the edge of the band experience localization, whereas (at least some) eigenstates near the center of the band as shown in Fig. 14.6.

When the chemical potential is anywhere in the localized states, then at low enough temperature, the electrons cannot move at all. Although there are states at this energy, they are all localized and electrons cannot jump between them. Hence we expect in this case that the DC dissipitave conductance goes to zero. (For dissipitive conductance to occur, an electron has to be excited up to the next delocalized band.) The state remains incompressible for filling fractions even away from the precise integer value of ν.

What is not obvious is (a) that the Hall conductance should be precisely quantized, and (b) that we should have Hall conductance at all.

14.6 Laughlin's Quantization Argument

In 1981, shortly after von Klitzing's discovery of quantum Hall effect, Bob Laughlin[8] presented an argument as to why the Hall conductance must be precisely quantized. The argument relies on gauge invariance. We first need to present a key theorem which comes from gauge invariance.

14.6.1 Byers and Yang Theorem

Consider any system (made of electrons and protons and neutrons) with a hole cut in it, as in Fig. 14.7.

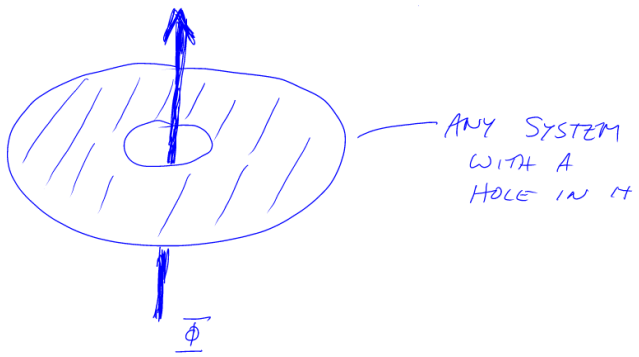


Figure 14.7: The Byers-Yang theorem states that threading any integer number of flux quanta through a hole in a system leaves the eigenspectrum unchanged.

Now put some magnetic flux Φ through the hole in such a way that the flux does not touch any piece of the system, but just goes through the hole. By the Aharanov-Bohm effect, the charged particles in the system cannot detect the flux if it is an integer multiple of the flux quantum ϕ_0. In fact the statement can be made stronger: The eigenspectrum of the system is precisely the same when an integer number of flux is inserted through the hole. This result is known as the Byers[9]-Yang[10] theorem (1961).

[8] Laughlin would later go on to win a Nobel prize for his explanation of *fractional* quantum Hall effect, which we will start discussing in the next chapter.

[9] Nina Byers was just starting as an assistant professor at UCLA when she proved this theorem. In the late 60s and early 70s she oscillated between Oxford (Somerville college) and UCLA, but eventually converged to UCLA. She told me personally that she regretted leaving Oxford. She passed away in 2014.

[10] Yang is C.N.Yang, who won a Nobel prize in 1957 along with T. D. Lee for his prediction of parity non-conservation of the weak interaction.

To prove this theorem we use gauge invariance. One is always free to make a gauge transformation

$$\mathbf{A}'(\mathbf{r}) = \mathbf{A}(\mathbf{r}) + (\hbar/e)\nabla\chi(\mathbf{r})$$
$$\Psi'(\mathbf{r}_1, \ldots \mathbf{r}_N) = \left[\prod_{j=1}^{N} e^{i\chi(\mathbf{r}_j)}\right] \Psi(\mathbf{r}_1, \ldots \mathbf{r}_N)$$

which leave the physical electromagentic field completely unchanged and changes the gauge of the wavefunction. The meaning of gauge invariance is that if we have a solution to the Schroedinger equation for Ψ and $\mathbf{A}$ at energy E, then we also have a solution at the same energy E for Ψ' and $\mathbf{A}'$.

When the physical geometry we are concerned with is non-simply connected, we can make gauge transforms which are non-single-valued, such as

$$\chi(\mathbf{r}) = m\theta(\mathbf{r})$$

wnere θ is the angle around the center. Making this gauge transform leaves the eigenspectrum of the system unchanged. However, the flux enclosed

$$\Phi' = \oint \mathbf{A}' \cdot \mathbf{dl} = \oint \mathbf{A} \cdot \mathbf{dl} + 2mh/e = \Phi + m\phi_0$$

has changed by an integer number of flux quanta.

14.6.2 Quantization of Hall Conductance

Laughlin's argument applys the Byers-Yang theorem to the Quantum Hall case. Consider a two dimensional electron system cut in an annulus[11] as shown in Fig. 14.8.

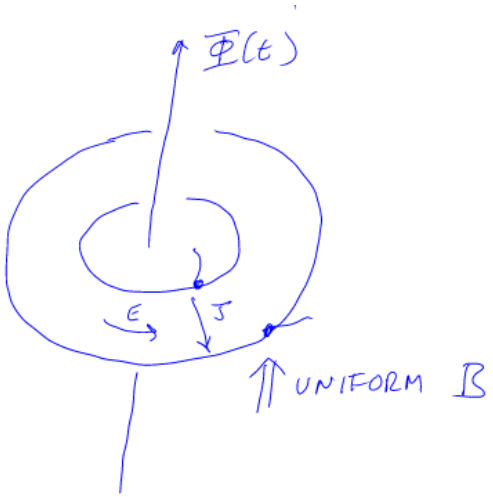


Figure 14.8: Insertion of Flux $\Phi(t)$ through the center of an annulus of two-dimensional electrons in a uniform magnetic field. Adiabatically increasing the flux creates an electric field in the annular direction which then, by the Hall conductivity, creates current in the radial direction.

Here we put the entire system in a uniform magnetic field (so that we have Landau levels) and we arrange such that the chemical potential is in the localized part of the band so that at low enough temperature the longitudinal (dissipitive) conductivity is zero.

[11] For studying current flow in magnetic fields, the annulus is knowni as "Corbino" geometry, after O. M. Corbino, who studied this in 1911.

We then adiabatically insert an additional flux $\Phi(t)$ through the center of the annulus and turn it on slowly from zero to one flux quantum. Due to the Faraday's law, an EMF is generated around the annulus

$$\mathcal{E} = -\frac{d\Phi}{dt} = \oint \mathbf{dl} \cdot \mathbf{E}$$

If there is a Hall conductance, G_H then this generates a radial current

$$J = G_H \mathcal{E}$$

As we slowly increase the flux by an amount $\Delta\Phi$ we have a total charge ΔQ moved from the inside to the outside of the annulus given by

$$\Delta Q = \int dt J(t) = G_H \int dt\, \mathcal{E}(t) = -G_H \int dt \frac{d\Phi(t)}{dt} = -G_H \Delta\Phi$$

Now the key to the argument is the Byers-Yang theorem. If we choose $\Delta\Phi = \phi_0$ a single flux quantum, then the final eigenstates of the system must be precisely the same as the initial eigenstates of the system. Since we have changed the system adiabatically (and there is a gap to excitations when the states at the chemical potential are localized due to disorder) the system must stay in the ground state[12] and the insertion of the flux quantum must take us from the ground state back to the very same ground state. The only thing that might have changed during this process is that an *integer* number p of electrons may have been transferred from the inside of the annulus to the outside. Thus we have

$$-pe = \Delta Q = -G_H \Delta\Phi = -G_H \phi_0 = -G_H (h/e)$$

Thus we obtain the quantized Hall conductance

$$G_H = p(e^2/h)$$

with p an integer!

Thus we see that the Hall conductance experiment is really some sort of "spectroscopy" to measure the charge on the electron! (hence the precision of the effect).

14.6.3 The Halperin Refinement

Although we have shown the the Hall conductance must be quantized, what we have not shown is that it must be nonzero! Afterall, since the chemical potential is in a localized band, it looks like electrons simply can't move at all.

A more careful argument was made by Halperin immediately after Laughlin's initial work. The key here is to think of a geometry where much of the system is free of disoder. In particular we consider the geometry shown in Fig. 14.9.

[12] There is a subtlely here. With disorder, there are actually low energy excitations, but they require very long range hops of localized electrons which cannot be made. So the system is "locally" gapped.

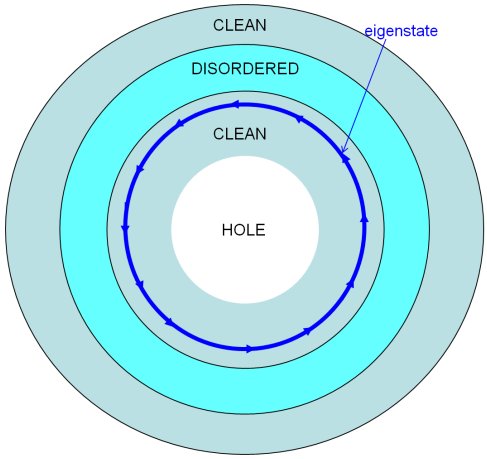


Figure 14.9: The Halperin geometry. The same as the Laughlin annulus geometry, except here we add disorder only in part of the annulus. We have also shown (dark blue) a single particle eigenstate in the clean region, which forms a circle (with a small gaussian cross-section).

Here, the disorder is confined to only part of the annulus, the inner-most and outer-most regions of the annulus being disorder-free. Within the clean regions we can solve for the eigenstates using symmetric gauge (this is a homework problem, but we will also discuss further in the next chapter). The eigenstates are indexed by their angular momentum m, and in the Lowest Landau level, for example, they are given by

$$\varphi_m \sim z^m e^{-|z|^2/(4\ell^2)}$$

where $z = x+iy$ is the complex representation of the position. A radial cut of one of these eigenstates gives a gaussian wavepacket[13] at radius $\ell\sqrt{2m}$— very similar to what we had in Landau gauge, but now these eigenstates are indexed by angular momenta instead of linear momenta, and they go around in circle instead of going straight.

Let us imagine the chemical potential above the middle of a Landau level (say above the middle of the lowest Landau level) until it sits in a localized piece (at least within the disordered region the wavefunctions are localized). Since this is above the middle of the Landau level, the Landau level is completely filled in the clean region.

Now, let us track what happens to the eigenstates as we change the flux through the hole. If the flux through the hole is an integer (in units of the flux quantum ϕ_0), then the angular momentum is also an integer. However, if the flux through the hole is an integer plus some fraction α, then the angular momentum quantum number must also be an integer plus α. Thus, as we adiabatically increase the flux by one flux quantum, we adiabatically turn each m eigenstate to $m+1$. Thus we are continuously pushing out electrons to the next further out radial wavefunction.

Now when we are in the disordered region of the annulus, we do not know any details of the shape of the eigenstates. All we know is that after insertion of a full flux quantum we must get back to the same many body eigenstate that we started with. However, we also know that an additional electron is being pushed into the disordered region from the clean region on the inside, whereas an electron is also being extracted into the clean region on the outside. Thus the disordered region must also convey exactly one electron (per Landau level) when a flux quantum is inserted adiabatically.

This argument pins down that the Hall conductance is not zero, but is h/e^2 times the number of Landau levels that are filled (in the clean regions).

[13] Just find the maximum of $|\psi_m|^2$.

Chapter 15

Introduction to Fractional Quantum Hall Effect

Having just determined that the quantum Hall effect is some sort of spectroscopy on the charge of the electron, it was particularly surprising in 1982 when Dan Tsui and Horst Stormer[1] discovered quantum Hall plateaus at fractional values of the filling fraction

$$\nu = p/q$$

with Hall resistance

$$R_H = \frac{h}{e^2}\frac{q}{p}$$

with p and q small integers. This effect is appropriately called the Fractional quantum Hall effect.

The first plateau observed was the $\nu = 1/3$ plateau[2] , but soon thereafter many more plateaus were discovered[3]. The Nobel prize for this discovery was awarded in 1998.

Given our prior gauge invariance argument that quantum Hall effect is measuring the charge of the electron — and that this is enforced by the principle of gauge invariance, it is hard to understand how the fractional effect can get around our prior calculation.

Two things must be true in order to have quantized Hall effect

> (a) Charge must fractionalize into quasiparticles with charge $e^* = e/q$, for example in the case of $\nu = 1/q$.
>
> (b) The ground state on an annulus must be degenerate, with q different ground states (in the case of $\nu = 1/q$) which cycle into each other by flux insertion through the annulus.

We should not lose sight of the fact that these things are surprising — even though the idea of degenerate ground states, and possibly even fractionalized charges, is something we have perhaps gotten used to in our studies of topological systems.

Given the Laughlin argument that inserting a flux though the annulus pumps an integer number of electrons from one side to the other, it is perhaps not surprising that fractional quantization of the Hall conductance must imply that a *fractional* charge has been pumped from one side of the annulus to the other (hence point (a) above). The way we get around the gauge invariance argument that implies the charge must be an integer is by having multiple degenerate ground states. In our argument for the Integer quantum hall effect we used adiabaticity, and the existence of a gap, to argue that we must stay in the ground state. However when there are multiple ground states (point

[1] Stormer had recently invented the idea of "modulation doping" semiconductors, which is a technique to obtain extremely clean two dimensional electron systems — a prerequisite for observing fractional quantum Hall effect.

[2] The legend is that Tsui very presciently looked at the data the moment it was taken and said "quarks!" realizing that the fractional plateau implied charge fractionalization!

[3] Over 60 different fractional quantum Hall plateaus have been discovered!

(b) above) we can only argue that we must always be in *some* ground state. Thus, for example, in the case of $\nu = 1/3$ where there are three ground states, the cycle of inserting flux is

$$\overset{\text{insert } \phi_0}{\longrightarrow} \quad |GS_1\rangle \quad \overset{\text{insert } \phi_0}{\longrightarrow} \quad |GS_2\rangle \quad \overset{\text{insert } \phi_0}{\longrightarrow} \quad |GS_3\rangle \quad \overset{\text{insert } \phi_0}{\longrightarrow} \quad |GS_1\rangle \quad \overset{\text{insert } \phi_0}{\longrightarrow}$$

where GS here means ground state. Each insertion of flux pumps $e^* = e/3$ charge from one side to the other. After three fractionally charged particles move from one side to the other, this amounts to a single electron being moved from one side to the other, and we return to exactly the same ground state as we started with.

So now we need only figure out how it is that this unusual situation of fractionalized charges, and multiple ground states (indeed, this situation of a topological quantum field theory!) comes about.

Want an incompressible state: Ignore disorder for now

We need to understand how we have an incompressible state when a Landau level is partially filled. As with the integer case, disorder will be important in allowing us to have plateaus of finite width, but the fundamental physics of the fracitonal quantum Hall effect comes from the fact that we have a gapped incompressible systems at a particular filling fraction. We can thus choose to consider a system free from disorder with the understanding that localization of excitations will be crucial to actually observe a plateau.

Why This is a Hard Problem: Massive Degeneracy

We restrict our attention to a clean system with a partially filled (say, 1/3 filled) Landau level. If there are N_e electrons in the system, there $3N_e$ available single electron orbitals in which to place these electrons. Thus in the absence of disorder, and in the absence of interaction, there are

$$\left(\begin{array}{c} 3N_e \\ N_e \end{array} \right) \sim (27/4)^{N_e}$$

multiparticle states to choose from — and all of these states have the same energy! In the thermodynamic limit this is an insanely enormous degeneracy[4]. This enormous degeneracy is broken by the interaction between the electrons, which will pick out a very small ground state manifold (in this case being just 3 degenerate ground states), and will leave the rest of this enormous Hilbert space with higher energy.

15.0.1 Our Model Hamiltonian

Since we are to neglect disorder, we can write the Hamiltonian for our system of interacting electrons as

$$H = \sum_i \frac{(\mathbf{p}_i + eA(\mathbf{r}_i))^2}{2m} + \sum_{i<j} V(\mathbf{r}_i - \mathbf{r}_j)$$

where the first term is just the kinetic energy of the electrons in the magnetic field, as discussed in Section 14.5, and the second term is the interaction beween the electrons, which we might take to be of $1/r$ Coulomb form, or perhaps a modified Coulomb form depending on the physical situation we are concerned with[5].

Now we have already analyzed the first term in this Hamiltonian back in Eq. 14.5, resulting in the structure of Landau levels. If we further assume that the cyclotron energy $\hbar\omega_c$ (the energy gap between Landau levels) is very large compared to the interacton energy scale V, then we can assume that there is very little effect of higher Landau levels — the interaction simply breaks the massive degeneracy of the partially filled Landau level without mixing in the higher Landau levels

[4]For example, if our system of size 1 square cm has a typically 10^{11} electrons in it, the number of degenerate states at $\nu = 1/3$ is roughly 10 to the 100 billion power! Way way way more than the number of atoms in the universe.

[5]For example, we could have a screened Coulomb potential if there are polarizable electrons nearby. The finite width of the quantum well also alters the effective Coulomb interaction.

(or putting holes in any completely filled Landau levels below the chemical potential). Another way to say this is that we are pursuing degenerate perturbation theory. The kinetic energy is completely determined (we just fill up Landau levels from the bottom up) and interaction only plays a role to break the degeneracy of the partially filled level.

The effective Hamiltonian is then just

$$H = \sum_{i<j} V(\mathbf{r}_i - \mathbf{r}_j)$$

where the Hilbert state is now restricted to a single partially filled Landau level. But here it might look like we are completely stuck. We have an enormously degenerate Hilbert space — and we have no small parameter for any sort of expansion.

Laughlin's insight was to simply guess the correct wavefunction for the system![6]. In order to describe this wavefunction we need to have a bit more elementary information about wavefunctions in a magnetic field (some of this is a homework problem!).

15.1 Landau Level Wavefunctions in Symmetric Gauge

We will now work in the symmetric gauge where the vector potential is written as

$$\mathbf{A} = \frac{1}{2}\mathbf{r} \times \mathbf{B}$$

where the magnetic field is perpendicular to the plane of the sample. (We can check that this gives $\nabla \times \mathbf{A} = \mathbf{B}$.

In this gauge, lowest Landau level wavefunctions (as mentioned before in section 14.6.3) take the form[7]

$$\varphi_m(z) = C_m z^m e^{-|z|^2/(4\ell^2)} \tag{15.1}$$

where

$$z = x + iy = re^{i\theta}$$

is the complex representation of the particle coordinate, $\ell = \sqrt{\hbar/eB}$ is the magnetic length, C_m is a normaliztion constant and here $m \geq 0$ is an integer. The most general lowest Landau level wavefunction for a single particle would be $f(z)$ times the gaussian factor for any analytic function f.

Note that the higher Landau level wavefunctions can all be obtained by application of a raising operator (which involve some prefactors of z^*) to the lowest Landau level wavefunctions. This algebra is discussed in a homework problem, so we will not belabor it here. A key point is that all Landau levels are effectively equivalent and one can exactly map any partially filled higher Landau level is equivalent to a partially filled lowest Landau level with an appropriately modified interaction. As such, we will focus exclusively on the lowest Landau level from here on.

Let us take a close look at the structure of the wavefunctions in Eq. 15.1. First we note that φ_m is an eigenstate of the angular momentum operator $\hat{L}$ (centered around the point $z = 0$)

$$\hat{L}\,\varphi_m = \hbar m\,\varphi_m$$

Secondly we should examine the spatial structure of φ_m. Writing $|\phi_m|^2 \sim r^{2m}\exp(-r^2/(2\ell^2))$ and differentiating with respect to r we find that the maximum of this function is at radius

$$r = \ell\sqrt{2m}$$

Thus the function roughly forms a gaussian ring at this radius. The area enclosed by this ring is $\pi r^2 = 2\pi m\ell^2 = m\phi_0/B$, which contains precisely m quanta of magentic flux.

[6] Decades of experience doing complicated perturbation theory led many people off on the wrong path — towards complicated calculations — when they should have been looking for something simple!

[7] We will ignore the spin degree of freedom as before.

15.1.1 What We Want in a Trial Wavefunction

In building a trial wavefunction for fractional quantum Hall effect, several rules will be important to follow

(1) **Analytic Wavefunction:** The wavefunction in the lowest Landau level should be comprised of single particle wavefunctions φ_m — that is, it must be a polynomial in z (with no z^*'s) times the gaussian factors. In other words we should have

$$\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N) = (\text{Polynomial in } z_1, \ldots z_N) \prod_{i=1}^{N} e^{-|z_i|^2/(4\ell^2)}$$

(2) **Homogeneous in Degree:** Since the Hamiltonian is rotationally invariant, we can expect that the eigenstates will be angular momentum eigenstates. Since the $\hat{L}$ operator counts powers of z, this means that the (Polynomial in $z_1, \ldots z_N$) part of the wavefunction must be homogeneous of degree.

(3) **Maximum Power of z_i is $N_\phi = N_e/\nu$:** Since the radius of the wavefunction is set by the exponent of z^m, the full radius of the quantum Hall droplet is given by the largest power of any z that occurs in the wavefunction. Since the area enclosed by the wavefunction should correspond to N_ϕ fluxes, this should be the maximum power.

(4) **Symmetry:** The wavefunction should be fully antisymmetric due to Fermi statistics, assuming we are considering fractional quantum Hall effect of electrons. It is actually very useful theoretically (and does not seem out of the question experimentally![8]) to consider fractional quantum Hall effect of bosons as well — in which case the wavefunction should be fully symmetric.

Even given these conditions we still have an enormous freedom in what wavefunction we might write down. In principle this wavefunction should depend on the particular interaction $V(r)$ that we put in our Hamiltonian. The miracle here is that, in fact, the details of the interaction often do not matter that much!

15.2 Laughlin's Ansatz

Laughlin simply guessed that a good wavefunction would be of the form[9]

$$\Psi^{(m)}_{Laughlin} = \prod_{i<j} (z_i - z_j)^m \prod_{i=1}^{N} e^{-|z_i|^2/(4\ell^2)}$$

The proposed wavefunction is properly analytic and homogeneous in degree. The maximum power of the wavefunction is

$$N_\phi = m(N-1)$$

thus corresponding to a filling fraction

$$\nu = N/N_\phi \to 1/m \quad \text{in large } N \text{ limit}$$

And the wavefunction is properly antisymmetric for m odd, and is symmetric for m even.

It is worth noting that for $m = 1$ the Laughlin wavefunction corresponds to a filled Landau level — that is, a single slater determinant filling all of the orbitals from $m = 0$ to $m = N_\phi = N - 1$. (This is a homework problem!)

[8] While no one has yet produced fractional quantum Hall effect of bosons, proposals for how to do this with cold atoms or interacting photons are plentiful, and it seems very likely that this will be achieved in the next few years.

[9] Note that this wavefunction is not normalized in any sense. The issue of normalization becomes important later.

It is also worth noting that the density of the Laughlin wavefunctio is completely constant in a disk up to its radius (and then the density falls quickly to zero). This constancy of density is proven by plasma analogy (which is another homework problem)[10].

Why should we think this wavefunction is particularly good? As two particles approach each other, the wavefunction vanishes as m powers. This means that the particles have low probability of coming close to each other — thus keeping the interaction energy low.

Being that the polynomial in each variable is of fixed total degree N_ϕ, the polynomial has a fixed number of analytic zeros. For the Laughlin wavefunction *all* of these zeros are on the positions of the other particles – thus the wavefunction arranges that the particles stay as far away from each other as possible in some sense.

15.2.1 Exact statements about Laughlin Wavefunction

It turns out that the Laughlin wavefunciton is actually the exact ground state of a special interparticle interaction[11].

Bosons at $\nu = 1/2$

Consider a system of bosons with the interparticle interaction given by[12]

$$V = V_0 \sum_{i<j} \delta(\mathbf{r}_i - \mathbf{r}_j)$$

with $V_0 > 0$. This is a non-negative definite interaction.

It is clear that the $\nu = 1/2$ Laughlin state of bosons $\Psi^{(m=2)}_{Laughlin}$ has zero energy for this interaction, since there is zero amplitude of any two particles coming to the same point. Further, however, the Laughlin state is the highest density wavefunction (lowest degree polynomial) that has this property[13]. For example, the Laughlin state times any polynomial is also a zero energy state of this interaction, but since it has been multiplied by a polynomial, the total degree of the wavefunction is higher, meaning the wavefunction extends to higher radius, making the system less dense. A schematic of the ground state energy as a function of filling fraction for this case is shown in Fig. 15.1.

[10] Roughly the story is as follows. The probability $|\Psi(z_1, \ldots, z_N)|$ of finding particles at position $z_1, \ldots, z_N$ can be phrased as a classical stat mech problem of a one-component 2d coulomb plasma in a background charge, by writing

$$|\Psi|^2 = e^{-\beta U(z_1,\ldots,z_N)}$$

with $\beta = 2/m$ and

$$U = -m^2 \sum_{i<j} \log(|z_i - z_j|) + \frac{m}{4} \sum_i |z_i|^2$$

where the first term is the coulomb interaction in 2d, and the second term is a background charge – which happens to be the charge associated with a uniform positve background (an easy thing to check using gauss's law). Assuming this plasma screens the background charge, it will be of uniform density up to a constant radius

[11] This was discovered by Haldane in 1983, then again by Trugman and Kivelson and also Pokrovski and Talapov in 1985.

[12] Actually this is a very realistic interaction for cold atom bosonic quantum Hall effect, should it be produced in the future.

[13] Although with some thought this fact seems obvious, proving it rigorously is tricky.

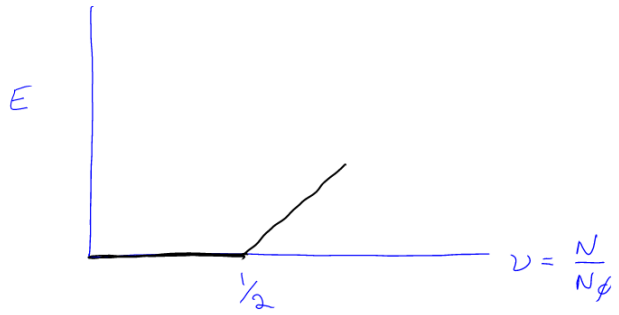


Figure 15.1: Schematic of the ground state energy as a function of filling fraction for bosons with delta function interaction.

The key point is that the ground state energy has a cusp, which means there is a jump in the chemical potential

$$\mu = \frac{\partial E}{\partial N}$$

This is precisely the same "incompressibility" as we have in the case of noninteracting electrons — where the chemical potential jumps between Landau levels! As in that case we presume that the presence of a cusp in the free energy, in the absence of disorder, will be enough to give us a plateau when disorder is added back in.

Now while we can easily show that there is a change of behavior at $\nu = 1/2$ in this plot, it is somewhat more difficult to be convincing that the slope coming from the right is finite — i.e., that the gap is actually finite. In order to do that, we would need to think about the elementary excitations – or resort to numerics.

Fermions at $\nu = 1/3$

The arguments given for bosons at $\nu = 1/2$ can be easily generalized to the case of fermions (i..e, electrons) at $\nu = 1/3$ (and more generally to any $\nu = 1/m$.) Obviously a δ-function interaction will no longer do the job, since for fermions Pauli exclusion prevents any two fermions from coming to the same point already. However, consider an interaction of the form

$$V = V_0 \sum_{i<j} \nabla^2 \delta(\mathbf{r}_i - \mathbf{r}_j)$$

Given a wavefunction $\Psi(r_1, \ldots, r_N)$ the interaction energy will be

$$E = \sum_{i<j} \int \mathbf{dr}_1 \ldots \mathbf{dr}_N \; |\Psi|^2 \;\; \nabla^2 \delta(\mathbf{r}_i - \mathbf{r}_j)$$

Writing

$$\Psi(\mathbf{dr}_1 \ldots \mathbf{dr}_N) = \phi(z_1 \ldots z_N) \prod_{i=1}^{N} e^{-|z_i|^2/(4\ell^2)} \tag{15.2}$$

with ϕ meaing the analytic polynomial part, for fermionic wavefunctions (that must vanish when $\mathbf{r}_i = \mathbf{r}_j$) the expression for the energy can be integrated by parts[14] using $\nabla^2 = 4\partial_z \partial_{z^*}$ to give

$$E = \sum_{i<j} \int \mathbf{dr}_1 \ldots \mathbf{dr}_N \; |\partial_{z_i} \phi|^2 \;\; \delta(\mathbf{r}_i - \mathbf{r}_j) \prod_{i=1}^{N} e^{-|z_i|^2/(2\ell^2)}$$

[14]Generally one would expect derivatives of the gaussian part as well when we integrate by parts. However, because the polynomial is antisymmetric, the derivitive must act on the polynomial part to prevent the wavefunction from vanishing when particle coordinates coincide.

Thus we have a non-negative definite interaction. Further, if the wavefunction vanishes as a single power when two particles come together, then $\partial_z\phi$ will be nonzero and we will get a postive result (Since $\partial_{z_i}(z_i - z_j)$ is nonzero). However, if the wavefunction vanishes as three powersr $\partial_z\phi$ will remain zero (since $\partial_{z_i}(z_i - z_j)^3$ goes to zero when $z_i = z_j$)[15].

Thus, entirely analously to the above case of $\nu = 1/2$ with the δ-function interaction, the Laughlin $m = 3$ ($\nu = 1/3$) wavefunction is the exact ground state (unique highest density zero energy wavefunction) of the $\nabla^2\delta$-function interaction. With similar ideas, one can construct interactions for which any $\nu = 1/m$ Laughlin wavefunction is exact.

15.2.2 Real Interactions

Obviously electrons do not interact via a $\nabla^2\delta$ interaction. They interact via a Coulomb interaction[16] What is perhaps surprising is that the Laughlin wavefunction is an almost perfect representation of the actual ground state. This statement comes from numerical tests. For example, for 9 electrons (on a spherical geometry to remove edge effects) the dimension of the fully symmetry reduced Hilbert space[17] is 84, and yet the Laughlin trial wavefunction has an overlap squared of .988 with the exact ground state of the Coulomb interaction. This is absurdly accurate! Energy of the Laughlin wavefunction differs from the energy of the exact Coulomb ground state by less than a part in two thousand[18].

15.3 Quasiparticles

The Laughlin quantum hall ground state is a uniform density fluid (we will actually show this as a homework problem). Density perturbations are made in discrete units of charge known as *quasiparticles*. Positively charged bumps of charge (opposite the charge of the electron) are known as *quasiholes* and negatively charged bumps of charge (same charge of the electron) are *quasielectrons.*

15.3.1 Quasiholes

For the quasiholes, it is fairly easy to guess their wavefunction (and indeed this was done by Laughlin). We start by considering adding a quasihole at postion $\mathbf{0}$. This leaves the system rotationally invariant. We guess the solution

$$\Psi_{qh}(\mathbf{0}) = \left[\prod_{i=1}^{N} z_i\right] \Psi_{Laughlin}$$

where $\mathbf{0}$ indicates we have put the quasihole at position $\mathbf{0}$. Here the degree of the polynomial is increased by one for every variable. So each filled orbital gets pushed out to the next orbital. This leaves precisely one empty orbtial open at positon $\mathbf{0}$, thus leaving a positive charge

$$e^* = \nu e$$

since filling fraction ν means on average a fraction ν of the orbitals are filled. So leaving the orbital at the center completely empty is a positive charge of $+\nu$.

Another way to think about the same wavefunction is to imagine adiabatically inserting a quantum of flux ϕ_0 at positon $\mathbf{0}$. Analogous to the Laughlin argument for integer quantum Hall effect, This creates an azimuthal EMF. Since the system has quantized hall conductance $\sigma_{xy} = \nu e^2/h$, the

[15] Note that by antisymmetry the wavefunction must go as an odd number of powers as two particle positions approach each other.

[16] In higher Landau levels, although the interaction is Coulomb, when the single Landau level problem is mapped down to a single partly filled *lowest* Landau level, the interaction gets modified – this mainly effects the short range behavior.

[17] The full Hilbert space is 45207 dimensional!

[18] I need to recheck this number.

total charge created is $\nu e = \sigma_{xy}\phi_0$. Then the full flux quantum can be gauged away leaving only the quasihole behind.

One can make quasiholes at any location w analogously,

$$\Psi_{qh}(w) = \left[\prod_{i=1}^{N}(z_i - w)\right]\Psi_{Laughlin}$$

although this is no longer an angular momentum eigenstate. We can similarly consider multiple quashioles the same way

$$\Psi_{qhs}(w_1, \ldots, w_M) = \left[\prod_{\alpha=1}^{M}\prod_{i=1}^{N}(z_i - w_\alpha)\right]\Psi_{Laughlin}$$

Several interesting comments at this point:

(1) While the z's are physical electron coordinates, the w parameters are simply parameters of the wavefunction and can be chosen and fixed to any value we like. The wavefunction $\Psi(w_1, \ldots w_M; z_1, \ldots z_N)$ is then the wavefunction of electrons z in the presence of quasiholes at fixed w positions.

(2) Note that the phase of the wavefunction wraps by 2π when any electron moves around the position of a quasihole.

(3) For the special ultra-short-range wavefunctions for which the Laughlin ground state is an exact zero energy eigenstate, then this Laughlin quasihole is also an exact zero energy eigenstate (albeit one with lower density than the ground state since a hole has been inserted). Take for example the case of $\nu = 1/2$. With a δ-function interaction, the energy is zero because no two particles come to the same point. Multiplying this wavefunction by any polynomial (as we have done to insert quasiholes) maintains this property and we still have a zero energy eigenstate. As for the Laughlin ground state, the quasihole is not exact for the Coulomb interaction, but is extremely accurate.

(4) At $\nu = 1/m$, if we insert m quasiholes at the same point w, then the wavefunction is just the same as if we were to have an electron e at the point w (although the electron is not there). Thus we expect that "fusing" m quasiholes together should precisely make an anti-electron (or a real hole).

15.3.2 Quasielectrons

The quasi-electron is a bump of *negative* charge (i.e, same charge as the electron). Unlike the case of quasiholes, there are no exact wavefunctions that we know of for quasi-electrons (not even for special short range interactions).

Whereas the quasi-hole increases the total degree of the polynomial wavefunction (thereby decreasing the density of the system) the quasi-electron should decrease the total degree of the wavefunction. Again, Laughlin made a very good guess of what the wavefunction for the quasi-electron should be. Considering an quasi-electron at the origin, we can write

$$\Psi_{qe}(\mathbf{0}) = \left(\left[\prod_{i=1}^{N}\frac{\partial}{\partial z_i}\right]\phi\right)\prod_{i=1}^{N}e^{-|z_i|^2/(4\ell^2)}$$

where as in Eq. 15.2 we have written the Laughlin wavefunction as the polynomial part ϕ times the gaussian factors. Obviously the derivative correctly reduces the degree of the polynomial by one in each varaible z, thus reducing the net angular momentum of each paricle by one. Each particle moves to lower radius by one orbital, thus giving a pile-up of charge of $e* = -e\nu$ at the origin.

In analogy to (but opposite that of) the quasihole, we might have looked for a quasi-electron where electrons accumulate a phase of -2π when an electron moves around the quasiparticle. One

might think of the operator z^*, but this operator does not live in the lowest Landau level. However, the projection of this operator to the lowet Landau level is given by

$$P_{LLL} z^* = 2\ell^2 \frac{\partial}{\partial z}$$

(This is a homework assignment!).

As mentioned above, the Laughlin quasi-electron is not exact for any known system. However, it is a fairly good trial wavefunction numerically for the Coulomb interaction. Note however, that other forms for the quasi-electron wavefunction have been found to be somewhat more accurate.

One can move the quasielectron to any position in a similar way as for quasiholes giving a wavefunction of the form

$$\Psi_{qes}(w) = \left(\left[\prod_{i=1}^{N} \left(2\ell^2 \frac{\partial}{\partial z_i} - w^* \right) \right] \phi \right) \prod_{i=1}^{N} e^{-|z_i|^2/(4\ell^2)}$$

15.3.3 Fractional Charge and Statistics?

The quasiparticles of the Laughlin state thus have fractional charge. One should not lose sight of how surprising this is — that particles can emerge that are a fraction of the "elementary" particles of the system. If we lived at very low energy, we would experience these particles as the fundamental particles of the system and would not know of the existence of the underlying electron.

Once one accepts fractionalized charge, it is perhaps not surprising to discover that they also have fractional statistics. Proving this statement is nontrivial, and we will do it in several ways. Note that since the quasiparticles are charged, moving them around in a magentic field incurs phases. We would like thus like to compare the phase of moving a particle in a loop versus moving a particle in a loop when another particle might be inside the loop, see fig. 15.2

Figure 15.2: To find the statistical phase, we compare moving a particle in a loop versus moving it in the same loop when another particle is inside the loop.

We shall perform this comparison next after we introduce Berry's phase, which is the effect which produces the statistical phase we are interested in.

15.4 Digression on Berry's Phase

The Berry phase[19] is one of the most fundamental ideas of modern physics. We recall the adiabatic theorem. If you start in an eigenstate and change a Hamiltonian sufficiently slowly, and there are no level crossings, then the system will just track the eigenstate as it slowly changes — i.e., it remains in the instantaneous eigenstate. However, during this process it takes a bit of thought to figure out what happens to the phase fo the wavefunction.

To see how this correction arises, let us consider a Hamiltonian $H(\mathbf{R})$ which is a function of some general parameters which we will summarize as the vector $\mathbf{R}$. In our case these parameters are

[19] Berry's work on Berry Phase in 1984 had a number of precursors, most prominentaly the work of Pancharatnam in 1956.

going to represent the quasiparticle position — we will insert this information inot the Hamiltonian by having some trapping potential which induces the quasiparticle at the point $\mathbf{R}$ and we can then move around the trapping potential in order to move the particle. Let us write the instantaneous (here normalized!) eigenstate as $|\psi(\mathbf{R})\rangle$. So we have

$$H(\mathbf{R})|\psi(\mathbf{R})\rangle = E(\mathbf{R})|\psi(\mathbf{R})\rangle$$

Now let us write the full, time dependent wavefucntion as

$$|\Psi(t)\rangle = e^{i\gamma(t)}\,|\psi(\mathbf{R}(t))\rangle$$

so we are allowing for an additional phase out front of the instantaneous eigenstate. The time dependent Schroedinger equation is

$$i\hbar\frac{\partial}{\partial t}|\Psi(t)\rangle = H(\mathbf{R}(t))|\Psi(t)\rangle$$
$$\left[-\hbar\dot{\gamma} + i\hbar\frac{\partial}{\partial t}\right]|\psi(\mathbf{R}(t))\rangle = E(\mathbf{R}(t))|\psi(\mathbf{R}(t))\rangle$$

Projecting this equation onto the bra $\langle\psi(\mathbf{R})|$ we obtain

$$\dot{\gamma} = -E(\mathbf{R}(t))/\hbar - i\left\langle \psi(\mathbf{R}(t))\left|\frac{\partial}{\partial t}\right|\psi(\mathbf{R}(t))\right\rangle$$

Integrating over some path $\mathbf{R}(t)$ from some initial time t_i to some final time t_f gives

$$\gamma(t_f) - \gamma(t_i) = -\frac{1}{\hbar}\int_{t_i}^{t_f} E(\mathbf{R}(t))dt \quad - i\int_{\mathbf{R}_i}^{\mathbf{R}_f}\mathbf{dR}\cdot\langle\psi(\mathbf{R})\,|\nabla_{\mathbf{R}}|\,\psi(\mathbf{R})\rangle$$

The first term is the expected dynamical phase — just accumulating a phase with time proportional to the energy. The second term on the right is the Berry phase contribution — a line integral along the particular path that $\mathbf{R}(t)$ takes. Note that this term depends *only* on the geometry of the path and not on how long one takes to move through this path. In this sense is it s a *geometric* phase.

15.5 Arovas-Schrieffer-Wilczek Calculation of Fractional Statistics

This section follows the approach of Arovas, Schrieffer and Wilczek[20].

Let us consider a $\nu = 1/m$ wavefunction for a quasihole

$$\Psi(w) = \mathcal{N}(|w|)\left[\prod_{i=1}^{N}(z_i - w)\right]\Psi^{(m)}_{Laughlin}$$

and we will imagine moving around the position w in a circle of constant radius as shown in the right of Fig. 15.2. Here we have inserted a normalization constant out front, which can be shown to be a function of radius only. (This is argued by plasma analogy, which is part of the homework). We will then parameterize[21] the position of the particle by the angle θ and $w = |w|e^{i\theta}$.

The Berry phase from moving the particle in a loop will then be

$$\Delta\gamma = -i\int_0^{2\pi} d\theta\ \langle\Psi(\theta)|\partial_\theta|\Psi(\theta)\rangle$$

[20] Wilczek won a Nobel for his work on assymptotic freedom. Schrieffer won a Nobel for his work on BCS theory of superconductivity. Arovas was a grad student at the time.

[21] On can choose a more general path for the particle but we will then need the detailed form of $\mathcal{N}(w)$. See the discussion below in section ***

where we have written $|\Psi(\theta)\rangle$ to mean $|\Psi(|w|e^{i\theta})\rangle$. We then have

$$\partial_\theta |\Psi(\theta)\rangle = \frac{\partial w}{\partial \theta}\left(\sum_i \frac{-1}{z_i - w}\right)|\Psi(\theta)\rangle$$

Thus we have

$$\langle\Psi(\theta)|\partial_\theta|\Psi(\theta)\rangle = \frac{\partial w}{\partial \theta}\sum_i \left\langle \Psi(\theta)\left|\frac{-1}{z_i - w}\right|\Psi(\theta)\right\rangle$$

Thus from taking w around in a circle we obtain the Berry phase[22]

$$\begin{aligned}\Delta\gamma &= -i\oint d\theta\,\langle\Psi(\theta)|\partial_\theta|\Psi(\theta)\rangle \\ &= -i\oint dw \sum_i \left\langle \Psi(w)\left|\frac{-1}{z_i - w}\right|\Psi(w)\right\rangle\end{aligned}$$

Now the integral around the loop of $1/(z-w)$ accumulates $2\pi i$ if and only if z_i is inside the loop. Thus we obtain the phase

$$\begin{aligned}\Delta\gamma &= 2\pi\ \langle\text{number of electrons in loop}\rangle \\ &= 2\pi(1/m)\Phi/\phi_0 = \gamma_{AB}\end{aligned}$$

where Φ is the flux enclosed by the loop and ϕ_0 is the flux quantum (and here we have used $\nu = 1/m$). This is precisely the expected Aharonov-Bohm phase that we should expect for moving a charge e/m around a flux Φ.

Now we consider putting another quasiparticle in the center of the loop as shown in the left of Fig. 15.2. Using a normalization factor that is again a function of $|w|$ only, the same calculation holds, but now the number of electrons enclosed has changed by one quasiparticle charge e/m. Thus the phase is now

$$\Delta\gamma = \gamma_{AB} + \gamma_{statistical}$$

where te additional phase for having gone around another quasihole is given by

$$\gamma_{statistical} = 2\pi/m$$

or in other words we have fractional statistics! For example, for the Laughlin state at $\nu = 1/2$, we have semionic statistics.

A more detailed version of this calculation (we will do this below) shows that the path of the particle does not matter —- the total phase is always the Aharanov-Bohm phase for taking a particle around flux, added to the statiscal phase of taking it around another quasiparticle.

Comment on the Fusion/Braiding Rules, and Chern-Simons theory

For the $\nu = 1/m$ Laughlin state thus have a situation where the elementary quasi-holes have statistics $\theta = 2\pi/m$. We can assume that their antiparticles will have the same statistics (both opposite "charge" and "flux" in a flux-charge model). We also have that the fusion of m elementary quasi-electrons or quasi-holes forms an an electron or anti-electron.

In the case where m is even, the underlying "electron" is a boson, in which case we can think of this electron as being identical to the vacuum — it has trivial braiding with all particles and it is essentially condensed into the ground state as some sort of background superfluid. Thus we have a simple anyon theory with m particle types.

On the other hand, when m is odd, we have the situation (discussed in our "charge-flux composite" section) where the fusion of m elementary anyons forms a fermion — and so there are actually $2m$ particle types — the fermion full-braids trivially with everything, but has fermionic statistics with itself. This situtation is "non-modular" — it does not have as many ground states as it has particle types. There are only m ground states, despite $2m$ particle types.

[22]The way this is written it is obviously a bit nonsense. Please fix it

15.6 Gauge Choice and Monodromy

The Lauglin wavefunction with M quasiholes takes the form

$$\Psi(w_1,\ldots,w_M;z_1,\ldots,z_N)=\mathcal{N}(w_1,\ldots,w_N)\left[\prod_{\alpha=1}^{M}\prod_{i=1}^{N}(z_i-w_\alpha)\right]\Psi^{(m)}_{Laughlin}(z_1,\ldots,z_N) \tag{15.3}$$

where $\mathcal{N}$ is a normalizing factor which can be thought of a an effective wavefunction for the quasiholes.

By using a plasma analogy (this is a homework assignment) we find that the normalization must be of the form

$$|\mathcal{N}(w_1,\ldots,w_M)|=C\prod_{\alpha<\beta}|w_\alpha-w_\beta|^{1/m}\prod_{\alpha=1}^{M}e^{-|w_\alpha|^2/(4\ell^{*2})}$$

where C is some constant and

$$\ell^*=\sqrt{\frac{\hbar}{e^*B}}$$

is the effective magnetic length for particle of charge $e^*=e/m$. This choice of normalization assures that

$$\langle\Psi(w_1,\ldots,w_M)|\Psi(w_1,\ldots,w_M)\rangle$$

is independent of the position of the quasiholes.

Now, we can choose the phase of the factor $\mathcal{N}$ arbitrarily – this is essentially a gauge choice. In the above Arovas, Schrieffer, Wilczek calculation above, we chose the phase to be real. However, this is just a convention. An intersting different convention is to choose

$$\mathcal{N}(w_1,\ldots,w_N)=C\prod_{\alpha<\beta}(w_\alpha-w_\beta)^{1/m}\prod_{\alpha=1}^{M}e^{-|w_\alpha|^2/(4\ell^{*2})} \tag{15.4}$$

which is known as holomorphic or "fractional statistics" gauge – here the fractional statistics of the quasiparticles are put explicitly into the wavefunction! Note here that this function is not single valued in the w-coordinates. In this gauge, we see that the wavefunction has branch cuts and can be thought of as having Riemann sheets. This may look problematic, but it is not. While a wavefunction must be single-valued in the physical electron coordinates, the w's are just parameters of the wavefunction, and we are allowed to choose wavefunctions the phase conventions in any way we like – even in non-single-valued ways as we have done here.

What we would want to confirm is that the physical phase accumulated in moving one quasihole around another is independent of our gauge choice. To this end we note that the total phase accumulated can be decomposed into two pieces, the so-called *monodromy* and the Berry phase. The monodromy is the phase explicitly accumulated by the wavefunction when one coordinate is moved around another.

$$\text{Total Phase} = \text{Monodromy} + \text{Berry Phase}$$

In the above Arovas-Schrieffer-Wilczek calculation, we chose the phase of the normalization to be everywhere real. So there is no monodromy — no, explicit phase as we move one particle around another. However, in fractional statistics gauge we see a phase of $2\pi/m$ for each particle which travels counterclockwise around another. In both gauges the total phase should be the same, so in the holomorphic gauge, the statistical part of the phase should be absent. Let us see how this happens

15.6.1 Fractional Statistics Calculation: Redux

Let us consider the case of two quasi-holes and repeat the argument of Arovas-Schrieffer-Wilczek but in holomorphic gauge. Putting one quasihole at postition w and another at position w' the

wavefunction is

$$\Psi(w) = C(w-w')^{1/m} e^{-(|w|^2+|w'|^2)/(4\ell^{*2})} \prod_i (z_i - w)(z_i - w') \prod_{i<j} (z_i - z_j) \prod_i e^{-|z_i|^2/(4\ell^2)}$$

with C chosen so Ψ is normalized independent of the quasihole coordinates[23]. Let us parameterize the path of a quasiparticle as $w(\tau)$. We can write the Berry phase as

$$\Delta\gamma = -i \oint d\tau \langle \Psi(\tau) | \partial_\tau | \Psi(\tau) \rangle$$

We write

$$\frac{\partial}{\partial \tau} = \frac{\partial w}{\partial \tau}\frac{\partial}{\partial w} + \frac{\partial w^*}{\partial \tau}\frac{\partial}{\partial w^*} \tag{15.5}$$

Now, because we are using holomorphic gauge of the wavefunction the $\partial/\partial w^*$ only hits the gaussian factor, so we have

$$\langle \Psi(w) | \partial_{w*} | \Psi(w) \rangle = -\frac{w}{4\ell^{*2}} \langle \Psi(w) | \Psi(w) \rangle = -\frac{w}{4\ell^{*2}}$$

To evaluate the derivative ∂/∂_w we integrate by parts so that it acts on the bra rather than the ket. Now since the bra is completely antiholomorphic in w except the gaussian, the derivative acts only on the gaussian again to give

$$\begin{aligned} \langle \Psi(w) | \partial_w | \Psi(w) \rangle &= \partial_w \left[\langle \Psi(w) | \Psi(w) \rangle \right] - \left[\partial_w \langle \Psi(w) | \right] | \Psi(w) \rangle \\ &= \frac{w^*}{4\ell^{*2}} \langle \Psi(w) | \Psi(w) \rangle = \frac{w^*}{4\ell^{*2}} \end{aligned}$$

Note that the derivative on $\langle \Psi | \Psi \rangle$ here is zero because the wavefunction is assumed normalized to unity for every value of w.

We then have the Berry phase given by

$$\Delta\gamma = -i \oint d\tau \langle \Psi(\tau) | \partial_\tau | \Psi(\tau) \rangle = -i \frac{1}{4\ell^{*2}} \oint (dw w^* - dw^* w)$$

where we have used Eq. 15.5. We now use the complex version of Stokes theorem[24] to obtain

$$\Delta\gamma = \frac{\text{Area}}{\ell^{*2}} = 2\pi(1/m)\Phi/\phi_0$$

which is the Aharanov-Bohm phase corresponding to the flux enclosed in the path – without giving the fractional statistical phase which has now been moved to the monodromy!

The key point here, which we emphasize, is that if we work with normalized holomorphic wavefunctions (i.e., holomorphic gauge), then the fractional statitics are fully explicit in the monodromy of the wavefunction — we can read the statistics off from the wavefunction without doing any work!

15.7 Appendix: Building an Effective (Chern-Simons) Field Theory

We can consider writing an effective field theory for this $\nu = 1/m$ quantum Hall system. First let us think about how it responds to an externally applied electromagnetic field. It should have its

[23] Strictly speaking the wavefunction is normalized in this form only if w and w' are not too close together — keeping them a few magnetic lengths apart is sufficient. This all comes from the plasma analogy calculation.

[24] The complex version of Stokes is as follows. Using $w = x + iy$

$$\int_{\partial A} (F dw - G dw^*) = -2i \int_A (\partial_{w^*} F + \partial_w G) dx dy$$

density locked to the magnetic field, so we should have a change of electron density (In this section we set $\hbar = e = 1$ for simplicity)

$$\delta n = j^0 = \frac{1}{2\pi m}\delta B$$

Similarly we should expect a quantized Hall conductance, here with j being the current of electrons

$$j^i = -\frac{1}{2\pi m}\epsilon^{ij}E_j$$

Both of these can be summarized as the response to a perturbing vector potential

$$j^\mu = \frac{-1}{2\pi m}\epsilon^{\mu\nu\lambda}\partial_\nu \delta A_\lambda \tag{15.6}$$

We must, of course have charge conservation as well. This is easy to enforce by writing the current in the form

$$j^\mu = \frac{1}{2\pi}\epsilon^{\mu\nu\lambda}\partial_\nu a_\lambda \tag{15.7}$$

which then automatically satisfies

$$\partial_\mu j^\mu = 0$$

In this language, the effective Lagrangian that produces Eq. 15.6 as an equation of motion is then

$$\mathcal{L} = \frac{-m}{4\pi}\epsilon^{\mu\nu\lambda}a_\mu\partial_\nu a_\lambda + \frac{1}{2\pi}\epsilon^{\mu\nu\lambda}A_\mu\partial_\nu a_\lambda + j_q^\mu a_\mu$$

where j_q is the quasiparticle current. Note that without the A_μ term, this is the same Chern-Simons theory we used for describing fractional statistics particles (now the quasiparticles).

To see the coupling to the external vector potential, note that the general (noether) current associcated with the local gauge symmetry will be

$$j^\mu = \frac{\partial \mathcal{L}}{\partial A^\mu}$$

which matches the expression from Eq. 15.7. By differentiating the Lagrangian with respect to a_μ we generate the equations of motion Eq. 15.6.

More here

15.8 Appendix: Quantum Hall Hierarchy

Good reference is https://arxiv.org/abs/1601.01697

Shortly after the discovery of the Laughlin $\nu = 1/3$ state additional fractional quantum Hall plateaus were discovered at filling fractions such as $\nu = 2/3, 2/5, 3/7$ and so forth. By now over 60 different plateaus have been observed in experiment!

The Laughlin theory only describes filling fractions $\nu = 1/m$ but it contains in it the right ideas to build possible theories for many of these fractions.

There are several approaches to building a hierarchy of quantum Hall states, however perhaps the most intuition comes from the original approaches by Haldane and Halperin in 1983.

The general idea is to begin with a Laughlin wavefunction for N electrons with coordinates z_i for $\nu = 1/m$ then change the magnetic field to add a large number M of quasiparticles (say in the form of 15.3, in the case of quasiholes) at coordinates w_α. Thus our wavefunction we write as

$$\Psi(w_1, \ldots w_M; z_1, \ldots z_N)$$

as written in Eq. 15.3. We then write a *pseudowavefunction* to describe some dynamics of the quasiholes which we write as

$$\phi(w_1, \ldots, w_M)$$

An electron wavefunction is generated by integrating out the quasihole coordinates. Thus we have

$$\tilde{\Psi}(z_1, \dots z_N) = \int \mathbf{dw}_1, \dots \mathbf{dw}_M \, \phi^*(w_1, \dots, w_M) \, \Psi(w_1, \dots w_M; z_1, \dots z_N)$$

The general idea of this scheme is that the pseudo-wavefunction can itself be of the form of a Laughlin wavefunction. In the original Laughlin argument we wrote down wavefunctions for both boson and fermion particles. Here, the particles w are anyons, so we need to write a slightly different form of a wavefunction. We expect

$$\phi(w_1, \dots, w_M) = \prod_{\alpha<\beta} (w_\alpha - w_\beta)^{\frac{1}{m}+p}$$

with p an even integer. The fractional power accounts for the fact that the anyon wavefunction must be multi-valued as one particle moves around another. The factor p is to include a "Laughlin" factor repelling these anyons from each other without further changing the statistics.

The condensation of these quasi-particles into a Laughlin state generates a wavefunction for the filling fraction

$$\nu = \frac{1}{m \pm 1/p}$$

with the $\pm$ corresponding to whether we are condensing quasiparticles or quasiholes. One can continue the argument starting with these new fractions and generating further daughter states and so forth. At the next level for example, we have

$$\nu = \frac{1}{m \pm \frac{1}{p \pm \frac{1}{q}}}$$

By repeating the procedure, any odd denominator fraction $\nu = p/q$ can be obtained.

Chapter 16

Quantum Hall Edges

The bulk of a quantum Hall system is gapped, but on a finite system there are always low energy modes on the edges. (This is always true for any *chiral* topological system. Although achiral systems can have fully gapped edges). Even though the bulk is incompressible, the shape of the edge can be deformed as suggested in Fig. 16.1.

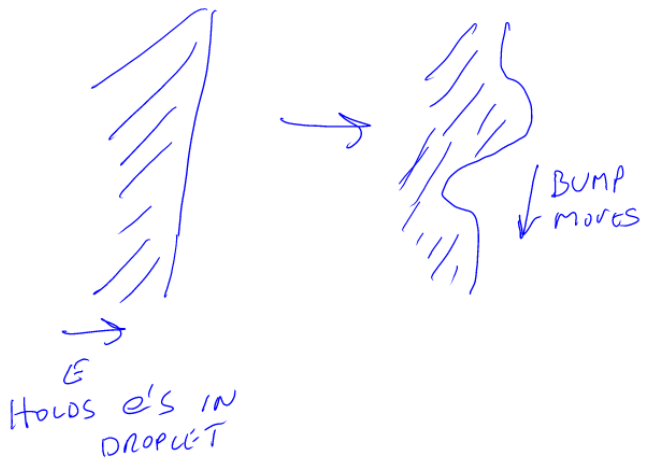


Figure 16.1: A deformation of the edge is a low energy edge excitation which moves along the edge due to $E \times B$ drift.

Now let us think about the dynamics of a bump on the edge. On the edge of the system we always have an electric field (this is the potential that holds the electrons in the system — otherwise they would just leak out!). Since we have $\mathbf{E} \times \mathbf{B}$, we expect a drift velocity for all the electrons on the edge. Thus we expect edge dynamics to be basically just movement of charge along the edge.

16.1 Landau Gauge Edge Picture for Integer Quantum Hall

(Can this section be moved to a chapter appendix?)

Recall in Landau gauge (See section 14.5) the wavefunctions are plane waves in the y direction, but are harmonic oscillator states in the x direction. We now impose an additional confining potential in the x direction near the edges of the system as shown in Fig. 16.2.

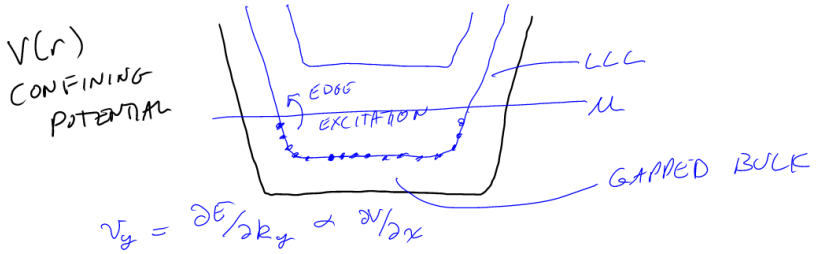


Figure 16.2: Low energy edge excitations

The addition of the confining potential $V(x)$ simply adds this potential to the 1-d schroedinger equation 14.2. If the confining potential is fairly smooth, it simply increases the energy of the eigenstates when the position $x = -k_y\ell^2$ gets near the edge of the system as shown in Fig. 16.2.

In the case of the integer quantum Hall effect, all of the eigenstates of some particular Landau level (the lowest Landau level in the figure) are filled within the bulk. At some point near the edge, the Landau level crosses through the chemical potential and this defines the position of the edge. Since the eigenstates are labeled by the quantum number k_y it is possible to create a low energy excitation by moving an electron from a filled state near the edge just below the chemical potential to an emtpy state near the edge just above the chemical potential. The excitation will have momentum $\hbar\Delta k_y$. [1] We thus have a 1-d system of fermions filled up to a chemical potential and they flow only in one direction along each edge — i.e., they are chiral fermions.

16.2 Parabolic Confinement

For studying fractional quantum Hall edge states, it is perhaps most useful to consider a parabolic confinement potential. Considering the simple particle Hamiltonian, and adding this confining potential to the kinetic energy we have

$$H_{confined} = H_0 + \gamma r^2$$

where H_0 is the single particle Hamiltonian in the asence of the confinement.

Since the confinement is rotationally symmetric, we can still classify all eigenstates by their angular momemtum quantum numbers. Using symmetric gauge we can still write the single particle

[1] The change in energy will be

$$\Delta E = \frac{\partial V}{\partial x}\Delta x = \frac{\partial V}{\partial x}\ell^2\Delta k_y$$

Thus the edge velocity is given by

$$v = \frac{1}{\hbar}\frac{\partial E}{\partial k} = \frac{1}{\hbar}\frac{\partial V}{\partial x}\ell^2$$

If the chemical potential along the one edge is raised by $\Delta\mu$, a range of k-states

$$\Delta k = \frac{\Delta\mu}{\ell^2\frac{\partial V}{\partial x}}$$

will be filled. Since the spacing between adjacent k states is $2\pi/L_y$ this corresponds to an increase in electrons per unit length along the edge of

$$\Delta n_{1d} = \frac{2\pi\Delta\mu}{\ell^2\frac{\partial V}{\partial x}}$$

These then carry a net 1d electron current density

$$j = -ev\Delta n_{1d} = -e(\frac{1}{\hbar}\frac{\partial V}{\partial x}\ell^2)\frac{2\pi\Delta\mu}{\ell^2\frac{\partial V}{\partial x}} = -(e/h)\Delta\mu$$

which is precisely the expected quantized Hall current flowing along the edge. ($\Delta\mu = -e\Delta V$).

eigenstates as[2]

$$\varphi_m \sim z^m e^{-|z|^2/\ell^2}$$

where m is the eigenvalue of the angular momentum[3] operator $\hat{L}$. Since the radius of these states is $r \approx \ell\sqrt{2m}$ it is not surprising that the confinement energy γr^2 of each eigenstate is proportional to m. We thus have

$$H_{confined} = H_0 + \alpha \hat{L}$$

for some constant α.

For integer filling, the edge excitations are very much like the edge excitations we discussed above in Landau gauge. A round quantum Hall droplet fills m states up to a chemical potential along the edge. One can add a small amount of angular momentum to the edge by exciting a filled state from an m just below the chemical potential to an empty state just above the chemical potential.

16.3 Edges of The Laughlin State

We now consider adding an interaction term so as to produce a fractional quantum Hall state. It is convenient to think about the limit where the cyclotron energy is huge (so we are restricted to the lowest Landau level), the interaction energy is large, so we have a very well formed quantum Hall state, and finally, the edge confinement is weak.

In particular if we choose to consider the special ultra-short range interaction potentials (such as δ function for bosons at $\nu = 1/2$) we still have the ground state given exactly by the Laughlin state

$$\Psi^{(m)}_{Laughlin} = \prod_{i<j}(z_i - z_j)^m \prod_{i=1}^{N} e^{-|z_i|^2/(4\ell^2)}$$

such that it has zero interaction energy. The angular momentum of the Laughlin ground state is just the total degree of the polynomial

$$L_{ground} = m\frac{N(N-1)}{2}$$

with confinement energy

$$E_{ground} = \alpha m\frac{N(N-1)}{2}$$

While the Laughlin state has zero interaction energy it is also the case that any polynomial times the Laughlin state also has zero interaction energy since multipying by a polynomial does not ruin the fact that the wavefunction vanishes as m or more powers as two particles approach each other. Thus we can consider all possible wavefunctions of the form

$$\Psi = (\text{Any Symmetric Polynomial})\,\Psi^{(m)}_{Laughlin}$$

where we insist that the polynomial is symmetric such that the symmetry of the wavefunction remains the same (i.e, antisymmetric for fermions and symmetric for bosons).

If the degree of the symmetric polynomial is ΔL, then we have

$$L = L_{ground} + \Delta L$$
$$E = E_{ground} + \alpha \Delta L$$

We can organize the possible excitations by their value of ΔL. We thus only need to ennumerate all possible symmetric polynomials that we can write in N variables of some given degree ΔL.

[2]Note tha t the parabolic confinement modifies the magnetic length.

[3]We drop the $\hbar$ from the angular momentum operator so its eigenvalues are just numbers.

We thus need some facts from the theory of symmetric polynomials. The symmetric polynomials on the N variables $z_1, \ldots, z_N$ form a so-called "ring" (this means you can add and multiply them). A set of generators for this ring is given by the functions

$$p_m = \sum_{i=1}^{N} z_i^m$$

This means that any symmetric function on N variables can be written as sums of products of these functions[4]. Thus it is extremely easy to count symmetric functions. Of degree 1, we have only p_1. At degree 2, we have p_1^2 and also p_2. Thus the vector space of symmetric polynomials (with real coefficients) is two dimensional. We can build a corresponding table as shown in Table 16.1.

$L - L_{ground}$	dimension	basis functions	Energy
1	1	p_1	α
2	2	$p_2, \quad p_1p_1$	2α
3	3	$p_3, \quad p_2p_1, \quad p_1p_1p_1$	3α
4	5	$p_4, \quad p_3p_1, \quad p_2p_1p_1, \quad p_1p_1p_1p_1$	4α
5	7	$p_5,\ p_4p_1,\ p_3p_2,\ p_3p_1p_1,\ p_2p_2p_1,\ p_2p_1p_1p_1, \quad p_1p_1p_1p_1p_1$	5α

Table 16.1: Table of Symmetric Polynomials

Thus the number of edge excitations at a given angular momentum follows a pattern, $1, 2, 3, 5, 7, \ldots$ with energy increasing linearly with the added angular momentum. Note that this result holds also for the $\nu = 1$ Laughlin state (i.e., for the integer quantum Hall effect), and matches the counting for excitations of a chiral fermion (try this exercise![5])

16.3.1 Edge Mode Field Theory: Chiral Boson

An equivalent description of the edge modes is given by the Hamiltonian

$$H = \sum_{m>0} (\alpha m) b_m^\dagger b_m$$

where the $b_m^\dagger$ are boson creation operators satisfying the usual commutations

$$[b_m, b_n^\dagger] = \delta_{nm}$$

and we think of these boson creation operators $b_m^\dagger$ as creating an elemetary excitation of angular momentum m on the ground state which we will call $|0\rangle$ for now. We can build a table describing all of the states in fock space of this Hamiltonian, ordered by their angular momentum as shown in Table 16.2. We see the fock space is precisely equivalent to the above table of polynomials. In fact the analogy is extremely precise. In the thermodynamic limit, up to a known normalization constant, application of $b_m^\dagger$ is precisely equivalent to multiplication of the wavefunction by p_m.

[4] In fact because the interaction Hamiltonian that we are studying is purely real when written in the φ_m basis, we can take the coefficients in the polynomials to be entirely real too.

[5] To get you started, consider filled states in a line filled up to the chemical potential. We can think of these as dots in a row. For example, let the ground state be

$$\ldots \bullet\bullet\bullet\bullet\bullet\bullet \circ\circ\circ\circ \ldots$$

where $\bullet$ means a filled single particle eigenstate and $\circ$ means empty. Now if we add one unit of (angular) momentum, we have the unique state

$$\ldots \bullet\bullet\bullet\bullet\bullet \circ \bullet \circ\circ\circ \ldots$$

adding two units can be done in two ways

$$\ldots \bullet\bullet\bullet\bullet\bullet \circ\circ \bullet \circ\circ \ldots$$

and

$$\ldots \bullet\bullet\bullet\bullet \circ \bullet\bullet \circ\circ\circ \ldots$$

thus starting the series $1, 2, 3, 5, 7 \ldots$.

$L - L_{ground}$	dimension	basis fock states				Energy
1	1	$b_1^\dagger\|0\rangle$				α
2	2	$b_2^\dagger\|0\rangle$,	$b_1^\dagger b_1^\dagger\|0\rangle$			2α
3	3	$b_3^\dagger\|0\rangle$,	$b_2^\dagger b_1^\dagger\|0\rangle$,	$b_1^\dagger b_1^\dagger b_1^\dagger\|0\rangle$		3α
4	5	$b_4^\dagger\|0\rangle$,	$b_3^\dagger b_1^\dagger\|0\rangle$,	$b_2^\dagger b_1^\dagger b_1^\dagger\|0\rangle$,	$b_1^\dagger b_1^\dagger b_1^\dagger b_1^\dagger\|0\rangle$	4α

Table 16.2: Fock Space for Chiral Bosons

These operators describe a *chiral* boson – chiral because they only have positive angular momentum $m > 0$ not negative angular momentum.[6]

16.4 Appendix: Edges and Chern-Simons theory

The existence of the edge theory could have been predicted from the effective Chern-Simons Lagrangian of the bulk. As mentioned previously, the Abelian Chern-Simons action is gauge invariant on a *closed* manifold. However, for a manifold with boundary, the action is not gauge invariant. This is what is known as an anomaly. The solution to this problem is that the action *becomes* gauge invariant only once it is added to an action for the low energy edge theory! We will not go through the detailed argument for this here.

[6] An *achiral* bose field on a circle requires both positive and negative angular momentum modes).

Chapter 17

Conformal Field Theory Approach to Fractional Quantum Hall Effect

In the last chapter we saw that we have an edge theory which is a chiral boson — a 1+1 dimensional dynamical theory. We can think of this theory as being a 2 dimensional cut out of a 3 dimensional space-time manifold. Now in a well-behaved topological theory, it should not matter too much how we cut our 3-dimensional space-time manifold. Thus we expect that the same chiral bose theory should somehow also be able to describe our 2+0 dimensional wavefunction. Since all chiral topological theories have gapless edges, this approach can be quite general.

1+1 dimensional gapless theories can all be described by conformal field theories (CFTs) possibly perturbed by irrelevant operators. And conformal field theories in 1+1 dimension are particularly powerful in that they are exactly solvable models, which can be used to describe either the dynamics of 1+1 dimensional systems or classical statistical mechanical models in 2 dimensions.

While we cannot provide a complete introduction to CFT here (see Ginsparg's lectures, Fendley's notes, or for a much more complete discussion, see the Big Yellow Book), it turns out that we need very little of the machinery to proceed. Furthermore, a large fraction of this machinery will look extremely familiar from our prior study of TQFTs. Indeed, there is an extremely intimite connection between CFTs and TQFTs — and much of what we know about TQFTs has grown out of the study of CFTs.

We wil begin by seeing how this works for the chiral boson, which is perhaps the simplest of all 1+1d CFTs. Below we will show how the scheme works in more detail in the context of quantum Hall physics, this approach, first described by Moore and Read, has been extremely influential in the development of TQFTs and their relationship to the quantum Hall effect.

17.1 The Chiral Boson and The Laughlin State

The simplest CFT is the bose theory in 1d. We can write a bose field as a function of x and $\tau = it$ as

$$\Phi(z, z^*)$$

where $z = x + i\tau$ and $z^* = x - i\tau$ which correspond to left and right-moving coordinates. The interesting thing about the field is that it can be decomposed cleanly into holomorophic (left moving) and antiholomorphic (right moving) pieces.

$$\Phi(z, z^*) = \phi(z) + \bar{\phi}(z^*)$$

If we are interested in a chiral theory we focus only on the holomorphic piece $\phi(z)$. As free bose fields, we can use Wick's theorem on the fields ϕ and all we need to know is the single two point

correlator[1]

$$\langle \phi(z)\phi(z')\rangle = -\log(z-z')$$

Note that we thik of this correlation function as a correlation in a 1+1d theory.

From this chiral ϕ operator we construct the so-called vertex operators

$$V_\alpha(z) =: e^{i\alpha\phi(z)} :$$

where : : means normal ordering[2] A straightforward exercise (assigned as homework!) using Wick's theorem then shows that

$$\langle V_{\alpha_1}(z_1)V_{\alpha_2}(z_2)\dots V_{\alpha_N}(z_N)\rangle = e^{-\sum_{i<j}\alpha_i\alpha_j\langle\phi(z_i)\phi(z_j)\rangle} = \prod_{i<j}(z_i-z_j)^{\alpha_i\alpha_j}$$

so long as

$$\sum_i \alpha_i = 0 \tag{17.1}$$

(otherwise the correlator vanishes).

17.1.1 Writing the Laughlin Wavefunction

We then define an "electron operator" to be

$$\psi_e(z) = V_{\sqrt{m}}(z)$$

This then enables us to write the holomorphic part of the Laughlin wavefunction as

$$\Psi^{(m)}_{Laughlin} = \langle \psi_e(z_1)\psi_e(z_2)\dots\psi_e(z_N)\hat{Q}\rangle = \prod_{i<j}(z_i-z_j)^m$$

Note that the index m has to be chosen as an integer such that the wavefunction is single valued in the electron coordinates. Note that here although the correlator means a 1+1d theory, we are constructing a wavefunction for a 2d system at fixed time.

Here, the operator $\hat{Q}$ can be chosen in two different ways. One possibility is to choose $\hat{Q} = V_{-N\sqrt{m}}(\infty)$, i.e., a neutralizing charge at infinity such that Eq. 17.1 is satisfied and the correlator does not vanish. This approach is often used if one is only concerned with keeping track of the holomorphic part of the wavefunction (which we often do). A more physical (but somewhat more complicated) approach is to smear this charge uniformly over the system. In this case, the neutralizing charge, almost magically, reproduces precisely the gaussian factors that we want![3].

[1] This comes from the achiral result

$$\langle \Phi(z,z^*)\Phi(z',z'^*)\rangle = -\log(|z-z'|^2)$$

To see where this comes from, it is easiest to think about a 2d classical model where the action is

$$S = (8\pi)^{-1}\int dxdy|\nabla\Phi|^2$$

With a partition function

$$Z = \int \mathcal{D}\Phi\ e^{-S[\Phi]}$$

It is then quite easy to calculate the correlator $\langle \Phi_k\Phi_{k'}\rangle = \delta_{k+k'}|k|^{-2}$. Fourier transforming this then gives the result.

[2] The usual understanding of normal ordering is that when we decompose a field into creation and annihilation operators, we can normal order by moving all the annihilation operators to the right. Another way to understand it is that when we expand the exponent $e^{i\alpha\phi(z)} = 1 + i\alpha\phi(z) + (i\alpha)^2\phi(z)\phi(z) + \dots$. There will be many terms where $\phi(z)$ occurs to some high power and that looks like a divergence because the correlator of two ϕ fields at the same position looks log divergent. Normal ordering is the same as throwing out these divergences.

[3] To see how this works, we divide the background charge into very small pieces (call them β) to obtain a correlator of the form

$$e^{m\sum_{i<j}\log(z_i-z_j)-\epsilon\sqrt{m}\sum_{i,\beta}\log(z_i-z_\beta)}$$

17.1.2 Quasiholes

Let us now look for quasihole operators. We can define another vertex operator

$$\psi_{qh}(w) = V_\beta(w)$$

and now insert this into the correlator as well to obtain

$$\begin{aligned}\Psi_{qh}(w) &= \langle \psi_{qh}(w)\psi_e(z_1)\psi_e(z_2)\dots\psi_e(z_N)\hat{Q}\rangle \\ &= \left[\prod_i (z_i - w)^{\beta\sqrt{m}}\right]\Psi^{(m)}_{Laughlin}\end{aligned} \tag{17.2}$$

Since we must insist that the wavefunction is single valued in the z coordinates, thus we must choose

$$\beta = p/\sqrt{m}$$

for some positive integer p, where the minimally charged quasiparticle is then obviously $p = 1$. (Negative p is not allowed as it would create poles in the wavefunction).

Further, using this value of the the charge β, along with the smeared out background charge, we correctly obtain the normalizing gaussian factor for the quasiparticle

$$e^{-|w|^2/(4m\ell^2)}$$

This is the correct gaussian factor, with an exponent $1/m$ times as big because the charge $V_{1/\sqrt{m}}$ is $1/m$ times as big as that of the electron charge $V_{\sqrt{m}}$.

If we are now to add multiple quasiholes, we obtain the wavefunction

$$\begin{aligned}\Psi(w_1,\dots,w_M) &= \langle \psi_{qh}(w_1)\dots\psi_{qh}(w_M)\,\psi_e(z_1)\dots\psi_e(z_N)\,Q\rangle \\ &= C\prod_{\alpha<\beta}(w_\alpha - w_\beta)^{1/m}\prod_{\alpha=1}^{M} e^{-|w_\alpha|^2/(4\ell^{*2})}\left[\prod_{\alpha=1}^{M}\prod_{i=1}^{N}(z_i - w_\alpha)\right]\Psi^{(m)}_{Laughlin}\end{aligned} \tag{17.3}$$

which is properly normalized

$$\langle\Psi(w_1,\dots w_M)|\Psi(w_1,\dots w_M)\rangle = \text{Constant}$$

and is in holomorphic gauge. As discussed previously in chapter *** with a normalized holomorphic wavefunction we can simply read off the fractional statistics as the explicit monodromy.

Note that we can consider fusion of several quasiparticles

$$V_{1/\sqrt{m}} \times V_{1/\sqrt{m}} \to V_{2/\sqrt{m}}$$

Fusion of m of these elementary quasiholes produces precisely one electron operator $V_{\sqrt{m}}$. Since the electrons are "condensed" into the ground state, we view them as being essentially the identiy operator (at least in the case of m even, which means we are considering a Laughlin state of bosons). Thus there are m species of particle in this theory. In the case of m odd, we run into the situation mentioned in chapter *** where the electron is a fermion, so really there are $2m$ species of particles in the theory.

---

the term with ϵ^2 we throw away as we will take the limit of small ϵ. Now here we realize that we are going to have a problem with branch cuts around these small charges – which we can do if we work in a funny gauge. Changing gauge to get rid of the branch cuts we then get only the real part of the second term. The second term is then of the form

$$\sum_{i,\beta}\log(|z_i - z_\beta|) \to \int d^2r\log(|z - r|)$$

where we have taken the limit of increasing number of smaller and smaller charges. We define this integral to be $f(z)$. It is then easy to check that $f(z) \sim |z|^2$ which is most easily done by taking $\nabla^2 f(z)$ and noting that log is the coulomb potential in 2d so Gauss's law just gives the total charge enclosed. Thus we obtain $e^{-|z|^2}$ as desired. A more careful calculation gives the constant correctly as well.

The idea is that by using conformal field theory vertex operators we automatically obtain normalized holomorphic wavefunctions and we can determine the statistics of quasiparticles straightforwarldy. This is a key feature of the Moore-Read approach. While there is no general proof that this will always be true, it appears to hold up in many key cases.

We hope now to generalize this construction by using more complicated conformal field theories. This then to generate more complicated fractional quantum Hall wavefunctions corresponding to more complcated TQFTs.

17.2 What We Need to Know About Conformal Field Theory

I can't possibly explain CFT in a few pages. (See the big yellow book. Ginsparg's lectures are nice for introduction. So are Fendley's notes).

But given what we already know about TQFTs many of the rules are going to seem very natural. Indeed, much of the math of TQFTs arose via CFTs.

CFTs are quantum theories in 1+1 dimension[4]. They are generically highly interacting theories, and most often it is impossible to write an explicit Lagrangian for the theory, but due to the special properties of being in 1+1 and having conformal invariance (guaranteed by being gapless in 1+1 d) these models are exactly solvable.

A particular CFT is defined by certain information known as conformal data, which basically mimics the defining features of a TQFT:

(1) There will be a finite set[5] of so-called **primary fields**, which we might call $\phi_i(z)$ (or we may use other notation). These are analogous to the particle types in a TQFT. Every CFT has an identity field often called I (which isn't really a function of position). Correlators of these fields

$$\langle \phi_{j_1}(z_1) \dots \phi_{j_N}(z_N) \rangle$$

are always holomorphic functions of the z arguments, although there may be branch cuts.

(2) Each primary field has a **scaling dimension**[6] or **conformal weight** or **conformal spin**, which we call h_i. The scaling dimension of I is $h_I = 0$. We have see these quantities before when we discussed twists in world lines. Often we will only be interested in h modulo 1, since the twist factor is $e^{2\pi i h}$. Each primary field has descendant fields which are like derivatives of the primary and they have scaling dimensions h_i plus an integer (we will typically not need these, but for example, $\partial_z \phi_i$ has scaling dimension $h_i + 1$).

(3) Fusion relations exist for these fields, which are associative and commutative

$$\phi_i \times \phi_j = \sum_k N_{ij}^k \phi_k$$

where fusion with the identity is trivial

$$I \times \phi_j = \phi_j$$

[4] We will restrict our attention to unitary CFTs so that these are well behaved 1+1 d theories. Although certain 2 dimensional stat mech models can be related to non-unitary CFTs, these do not correspond to well behaved TQFTs.

[5] A nonrational CFT may have an infinite number of particle types, but these are badly behaved and do not appear to correpsond to TQFTs.

[6] In CFT we have the powerful relation that if we make a coordinate transform $w(z)$ then any correlator of primary fields transforms as

$$\langle \phi_{i_1}(w_1) \dots \phi_{i_N}(w_N) \rangle = \left(\frac{\partial w_1}{\partial z_1} \right)^{-h_{i_1}} \left(\frac{\partial w_N}{\partial z_N} \right)^{-h_{i_N}} \langle \phi_{i_1}(z_1) \dots \phi_{i_N}(z_N) \rangle$$

However, we will not need this relationship anywhere for our discussion!

As with TQFTs, each particle type has a unique antiparticle. We will give a clearer meaning to these fusion relations in a moment when we discuss operator product expansion.

(4) The expectation of any correlator in the theory is zero unless all the fields inside the correlator fuse to the identity. For example, if we have a $\mathbb{Z}_3$ theory where it requires three ψ particles fuse to the identity, then we would have $\langle\psi(z)\psi(w)\rangle = 0$. We saw this law previously in the neutrality condition for the chiral boson. The expectation of the identity I is unity.

The fundamental theorem we need, which is beyond the simple analogy with TQFT is the idea of an **operator product expansion**. The idea is that if you take two field operators in a conformal field theory and you put them close together, the product of the two fields can be expanded as sum of resulting fields

$$\lim_{w\to z} \phi_i(w)\phi_j(z) = \sum_k C_{ij}^k (w-z)^{h_k-h_i-h_j}\phi_k(z) + \dots$$

Here the C_{ij}^k are coefficients which crucially are zero when N_{ij}^k is zero. In other words, when two fields are taken close together, the result looks like a sum of all the possible fusion products of these field. On the right hand side note that by looking at the scaling dimensions of the fields, we obtain explicit factors of $(w-z)$. The $\dots$ terms are terms that are smaller (less singular) than the terms shown and are made of descendant fields and higher powers of $(w-z)$. Crucially, no new types of branch cuts are introduced except those that differ by integers powers from (and are less singlar than) those we write explicitly.

The convenient thing about the operator product expansion (or "OPE") is that it can be used *inside* expectation values of a correlator. So for example

$$\lim_{w\to z} \langle\psi_a(w)\psi_b(z)\ \psi_c(y_1)\psi_d(y_2)\dots\psi_n(y_m)\rangle = \sum_k C_{ab}^k (w-z)^{h_k-h_a-h_b}\langle\psi_k(z)\ \psi_c(y_1)\psi_d(y_2)\dots\psi_n(y_m)\rangle$$

17.2.1 Example: Chiral Boson

The free boson vertex V_α has scaling dimension

$$h_\alpha = \frac{\alpha^2}{2}$$

. The fusion rules are

$$V_\alpha V_\beta = V_{\alpha+\beta}$$

corresponding to the simple addition of "charges". The resulting operator product expansion is then

$$V_\alpha(w)V_\beta(z) \sim (w-z)^{\alpha\beta}V_{\alpha+\beta}(z)$$

where we have used the notation $\sim$ to mean in the limit where w goes to z, and where the exponent is here given as

$$h_{\alpha+\beta} - h_\alpha - h_\beta = \frac{(\alpha+\beta)^2}{2} - \frac{\alpha^2}{2} - \frac{\beta^2}{2} = \alpha\beta$$

17.2.2 Example: Ising CFT

The Ising CFT is actually the CFT corresponding to a 1+1 d fermion, so it is particularly simple. The theory has three fields, I, σ, ψ with scaling dimensions

$$h_I = 0$$
$$h_\sigma = 1/16$$
$$h_\psi = 1/2$$

The fact that $h_\psi = 1/2$ is an indication that it is a fermion. The nontrivial fusion rules

$$\psi \times \psi = I$$
$$\psi \times \sigma = \sigma$$
$$\sigma \times \sigma = I + \psi$$

As in the case of TQFTs, it is the multiple terms on the right hand side that make a theory nonabelian.

We can write the operator product expansion

$$\psi(w)\psi(z) \sim (w-z)^{h_I - h_\psi - h_\psi} I + \dots$$
$$\sim \frac{I}{w-z} + \dots$$

The antisymmetry on the right hand side is precisely the behavior one should expect from fermions. It is crucial to note that within the ... all terms are similarly antisymmetric (and are less singular). Similarly, we have

$$\psi(w)\sigma(z) \sim (w-z)^{h_\sigma - h_\sigma - h_\psi}\ \sigma(z) + \dots$$
$$\sim (w-z)^{-1/2}\ \sigma(z) + \dots$$

where again the ... indicates terms which have the same branch cut structure but are less singular . In other words, wrapping w around z should incur a minus sign for all terms on the right.

Finally we have the most interesing OPE[7]

$$\sigma(w)\sigma(z) \sim C^I_{\sigma\sigma}(w-z)^{-1/8} I + C^\psi_{\sigma\sigma}(w-z)^{3/8}\psi(z) + \dots$$

where all terms in the ... must have branch cuts that match one of the two leading terms.

Let us consider calculating a correlator,

$$\lim_{w\to z} \langle \sigma(w)\sigma(z)\rangle$$

Since from rule (4) above, the two fields must fuse to the identity, we must choose the identity fusion channel only from the OPE. We then obtain

$$\lim_{w\to z} \langle \sigma(w)\sigma(z)\rangle \sim (w-z)^{-1/8} \tag{17.4}$$

On the other hand, calculating

$$\lim_{w\to z} \langle \sigma(w)\sigma(z)\psi(y)\rangle$$

in order to fuse to the identity, we must choose the ψ fusion of the two σ fields such that this ψ can fuse with $\psi(y)$ to give the identity. We thus have

$$\lim_{w\to z} \langle \sigma(w)\sigma(z)\psi(y)\rangle \sim (w-z)^{3/8} \tag{17.5}$$

Similarly one can see that fusion of two sigmas in the presence of any even number of ψ fields will be similar to Eq. 17.4, whereas in the presence of any odd number of ψ fields it will be like Eq. 17.5.

Note that the Ising CFT is actually a free fermion theory, which means that we can use Wick's (fermionic) theorem for correlators of the ψ fermi fields with the added information that[8]

$$\langle \psi(z)\psi(w)\rangle = \frac{1}{z-w}$$

which is exactly true, not only in the OPE sense. However, we cannot use Wick's theorem on correlators of the σ fields which are sometimes known as "twist" fields — we can think of these as altering the boundary conditions

[7] Remember these exponents of 1/8 and 3/8 from the ising anyon homework problems?

[8] Insert footnote or appendix that derives this. See Yellow Book for now!

17.3 Quantum Hall Wavefunction Based on Ising CFT: The Moore-Read State

Let us try to build a quantum Hall wavefunction based on the Ising CFT. We must first choose a field which will represent our electron. One might guess that we should use the fermion field. However, when two ψ fields come together the correlator (and hence our wavefunction) diverges, so this cannot be acceptable. Instead, let us construct an electron field which is a combination of the ising ψ field and a chiral bose vertex V_α

$$\psi_e(z) = \psi(z)V_\alpha(z)$$

These two fields are from completely different 1+1d theories simply multiplied together.

We then look at the operator product expansion to see what happens when two electrons approach each other

$$\psi_e(z)\psi_e(w) \sim \frac{I}{z-w}(z-w)^{\alpha^2}V_{2\alpha}$$

So in order for this to not be singular, we must have α^2 be a positive integer. If we choose

$$\alpha^2 = m$$

with m odd we have an overall bosonic operator ($\psi_e(z)\psi_e(w) = \psi_e(w)\psi_e(z)$) whereas if we choose m even we have an overall fermionic operator. However, we cannot choose $m = 0$ since that leaves a singularity. Thus we have the electron operator of the form

$$V_e(z) = \psi(z)V_{\sqrt{m}}(z)$$

Using this propsoed electron operator we build the multi-particle wavefunction

$$\Psi = \langle \psi_e(z_1)\psi_e(z_2)\dots\psi_e(z_N)\, Q\rangle$$

where Q is the background charge for the bose field. Since the Ising and bose fields are completely seperate theories we can take the expectation for the bose field to give

$$\Psi = \langle \psi(z_1)\psi(z_2)\dots\psi(z_N)\rangle \prod_{i<j}(z_i - z_j)^m \prod_{i=1}^{N} e^{-|z_i|^2/(4\ell^2)}$$

where the correlator is now in the Ising theory alone.

Now the Ising correlator is just a correlator must be zero unless there are an even number of ψ fields (since we need them to fuse to the identity). If the number of fermi fields is indeed even, then we can use the fact that ψ is a free fermi field and we can invoke Wick's theorem to obtain

$$\langle \psi(z_1)\psi(z_2)\dots\psi(z_N)\rangle = \mathcal{A}\left[\frac{1}{z_1-z_2}\ \frac{1}{z_3-z_4}\dots\frac{1}{z_{N-1}-z_N}\right] \equiv \mathrm{Pf}\left(\frac{1}{z_i-z_j}\right) \tag{17.6}$$

Here $\mathcal{A}$ means antisymmetrize over all reordering of the z's. Here we have written the usual notation for this antisymmetrized sum Pf which stands for "Pfaffian"[9]. Thus we obtain the trial wavefunction based on the Ising CFT

$$\Psi = \mathrm{Pf}\left(\frac{1}{z_i-z_j}\right)\prod_{i<j}(z_i - z_j)^m \prod_{i=1}^{N} e^{-|z_i|^2/(4\ell^2)}$$

which is known as the Moore-Read wavefunction. For m odd this is a wavefunction for bososn and for m even it is a wavefunction for fermions. To figure out the filling fraction, we note that the Pfaffian prefactor only removes a single power in each variable. Thus the filling fraction is determined entirely by the power m, and is given (like Laughlin) by $\nu = 1/m$.

[9]Several interesting facts about the Pfaffian: A BCS wavefunction for a spinless superconductor can be written as $\mathrm{Pf}[g(\mathbf{r}_i - \mathbf{r}_j)]$ where g is the wavefunction for a pair of particles. Any antisymmetric matrix M_{ij} has a Pfaffian

$$\mathrm{Pf}[M] = \mathcal{A}[M_{12}M_{34}...]$$

. A useful fact is that $(\mathrm{Pf}[M])^2 = \mathrm{det}M$.

17.3.1 Some Exact Statements About the Moore-Read Wavefunction

For simplicity, let us consider the $m = 1$ case $\nu = 1$ for bosons, which is the easiest to think about analytically. The wavefunction does not vanish when two particles come to the same point, since the zero of the $(z_1 - z_2)$ can be canceled by the pole of the Pfaffian. However, it is easy to see that the wavefunction must vanish (quadratically) when *three* particles come to the same point (three factors from $(z - z)^1$ but then one factor in the denominator of the Pfaffian).

Note that, even were we to not have an explicit expression for the Moore-Read wavefunction we would still be able to use the operator product expansion to demonstrate that the wavefunction (for $m = 1$) must vanish quadratically when three particles come to the same point[10]

Analogous to the case of the Laughlin wavefunction, it turns out that the Moore-Read wavefunction (for $m = 1$) is the exact (highest density) zero energy ground state of a *three-body* delta function interacton

$$V = V_0 \sum_{i<j<k} \delta(\mathbf{r}_i - \mathbf{r}_j)\delta(\mathbf{r}_i - \mathbf{r}_k)$$

Similarly one can construct a potential for fermions such that the $\nu = 1/2$ Moore-Read state ($m = 2$) is the highest density zero energy state. This is quite analogous to what we did for the Laughlin state:

$$V = V_0 \sum_{i<j<k} [\nabla^2\delta(\mathbf{r}_i - \mathbf{r}_j)]\delta(\mathbf{r}_i - \mathbf{r}_k)$$

Non-Exact Statements

Although the Coulomb interaction looks nothing like the three body interaction for which the Moore-Read Pfaffian is exact, it turns out that $\nu = 1/2$ Moore-Read Pfaffian $m = 2$ is an extremely good trial state[11] for electrons at $\nu = 5/2$ interacting with the usual Coulomb interaction. This is very suggestive that the $\nu = 5/2$ is topologically equivalent to the Moore-Read Pfaffian wavefunction (i.e., they are in the same phase of matter)[12] Further, the most natural interaction for bosons, the simple two-body delta function interaction has a ground state at $\nu = 1$ which is extremely close to the Moore-Read $m = 2$ Pfaffian.

17.4 Quasiholes of the Moore-Read state

We now try to construct quasiholes for the Moore-Read Pfaffian wavefunction. As we did in Eq. 17.2, we want to write

$$\Psi_{qh}(w) = \langle \psi_{qh}(w)\psi_e(z_1)\psi_e(z_2)\dots\psi_e(z_N)\hat{Q}\rangle$$

but we need to figure out what the proper quasihole operator ψ_{qh} is.

[10] To see this, note that taking the first two particles to the same point gives

$$\lim_{z_2\to z_1} \psi_e(z_1)\psi_e(z_2) \sim IV_2(z_1)$$

Then fusing the third particle

$$\lim_{z_3\to z_1} \psi_e(z_3)V_2(z_1) \sim (z_3 - z_1)^2\psi V_3(z_1)$$

[11] Here we have used a mapping between Landau levels, that any partially filled higher Landau level can be mapped to a partially filled lowest Landau level at the price of modifying the inter-electron interaction. This mapping is exact to the extent that there is no Landau level mixing. I.e., that the spacing between Landau levels is very large.

[12] There is one slight glitch here. It turns out that with a half-filled Landau level, the wavefunction and its charge-conjugate (replace electrons by holes in the Landau level) are inequivalent! It is possible that the $\nu = 5/2$ state is actually in the phase of matter defined by the conjugate of the Moore-Read state. The breaking of the particle-hole symmetry is very weak and involves Landau-level mixing. This debate, which one of the two possibilities is realized in experiment, has not yet been fully resolved.

Laughlin Quasihole

One obvious thing to try would be to write a simple vertex operator

$$\psi_{qh}^{L}(w) = V_{\beta}(w)$$

Looking at the OPE we have

$$\psi_{qh}^{L}(w)\psi_e(z) \sim (w-z)^{\beta\sqrt{m}}\psi(z)$$

In order to have the correlator be single valued in z (i.e., no branch cuts) we must choose $\beta = p/\sqrt{m}$ for some integer p (the smallest quasihole of this type corresponding to $p = 1$ then). This generates the wavefunction

$$\begin{aligned}\Psi_{qh}^{L}(w) &= \langle \psi_{qh}^{L}(w)\, \psi_e(z_1)\psi_e(z_2)\dots\psi_e(z_N)\hat{Q}\rangle \\ &= \left[\prod_{i=1}^{N}(z_i - w)\right] \Psi_{Moore-Read}^{(m)}\end{aligned} \tag{17.7}$$

which is just a regular Laughlin quasihole factor. By the same arguments, the charge of this quasihole is $e^* = e\nu$.

Minimal quasihole

However, the Laughlin quasihole is not the minimal quasihole that can be made. Let us try using an operator from the Ising theory as part of the quasihole operator. Suppose

$$\psi_{qh}(w) = \sigma(w)V_{\beta}(w)$$

We then have the operator product expansion

$$\psi_{qh}(w)\psi_e(z) \sim [\sigma(w)\psi(z)]\left[V_{\beta}(w)V_{\sqrt{m}}(z)\right] \sim (w-z)^{-1/2}(w-z)^{\beta\sqrt{m}}$$

Thus in order for the wavefunction not to have any branch cuts for the physical electron z coordinates, we must choose $\beta = (p+1/2)/\sqrt{m}$ for $p \geq 0$, with the minimal quasihole corresponding to $p = 0$. Thus we have the minimal quasihole operator of the form

$$\psi_{qh}(w) = \sigma(w)V_{\frac{1}{2\sqrt{m}}}(w)$$

Note that when we consider correlators, by the general rule (4) from section 17.2, the operators must fuse to the identity in order to give a nonzero result. Thus, we must always have an even number of σ fields[13]. We thus consider the wavefunction of the form

$$\Psi_{qh}(w,w') = \langle \psi_{qh}(w)\psi_{qh}(w')\, \psi_e(z_1)\psi_e(z_2)\dots\psi_e(z_N)\hat{Q}\rangle \tag{17.8}$$

$$\begin{aligned}&= (w-w')^{\frac{1}{4m}}\ \ e^{-(|w|^2+|w'|^2)/4\ell^{*2}}\ \prod_{i=1}^{N}(w-z_i)^{1/2}(w'-z_i)^{1/2} \\ &\times\quad \langle \sigma(w)\sigma(w')\,\psi(z_1)\psi(z_2)\dots\psi(z_3)\rangle\ \ \prod_{i<j}(z_i-z_j)^m \prod_{i=1}^{N} e^{-|z_i|^2/(4\ell^2)}\end{aligned} \tag{17.9}$$

Several comments are in order here. First of all, from the first line of Eq. 17.9 it looks like there are branch cuts with respect to the z coordinates. However, these fractional powers are precisely canceled by branch cuts in the correlator on the second line. Secondly the charge of the quasihole is determined entirely by the power of the $(z-w)$ factor, since it tells us how much the electrons are pushed away from the hole. (The correlator does not give an extensive number of zeros (as in

[13] Like the Sith, they come in pairs.

Eq. 17.6). If the exponent of $(z-w)$ were one, this would be a regular Laughlin quasihole with charge $e\nu$, thus here we have a quasihole charge of

$$e^* = e\nu/2.$$

I.e., the Laughlin quasihole has fractionalized into two pieces! This charge is reflected in the effective magnetic length $\ell^* = \sqrt{\hbar/e^*B}$.

Note that this wavefunction is still an exact zero energy state of the special interaction discussed above for which the Moore-Read wavefunction is the exact highest density zero energy state (the wavefunction here is higher degree and thus less dense, as we would expect given that we have added quasiholes). We can demonstrate the current wavefunction is still zero energy by bringing together three electrons to the same point and examining how the wavefunction vanishes. Since this can be fully determined by the operator product expansion, it does not matter if we add quasiholes to the wavefucntion, the vanishing property of the wavefunction remains the same, and thus this is an exact zero energy state of the special interaction.

A Crucial Assumption

The wavefunction here is single valued in all electron coorrdinates (as it should be) and is holomorphic in all coordinates (all z's and w's) except for the gaussian exponential factors. In this holomorphic gauge, as discussed above, we can read off the fractional statistics of the quasiparticles *given the assumption that the wavefunction is properly normalized.* This is a crucial assumption and it is not a simple result of CFT, but always requires an assumption about some sort of plasma being in a screening phase — and often the mapping to a plasma is highly nontrivial[14]. Nonetheless, from extensive numerical work, it appears that physics is kind to us and that these wavefunctions do indeed come out to be properly normalized!

Fusion and Braiding of Two Quasiholes in Identity Channel (even number of electrons)

Let us assume that the number of electrons is even. In this case the two σ's of the quasiholes fuse to the identity as in Eq. 17.4. As the two quasiholes approach each other we then have[15]

$$\psi_{qh}(w)\psi_{qh}(w') \sim (w-w')^{\frac{1}{4m}-\frac{1}{8}}$$

where the $\frac{1}{4m}$ is written expliclty in the first line of Eq. 17.9 and the $-\frac{1}{8}$ is from the operator product expansion Eq. 17.4. Invoking now the crucial assumption that the wavefunctions are normalized, since they are obviously holomorphic, we simply read off the statistical phase (the monodromy) we get for wrapping one quasihole around another!

One might object that the operator product expansion only tells us the behavior of the correlator as w and w' come close to each other. However, we are guaranteed that there are no other branch cuts in the system — the only branch cut in the wavefunction for w is when it approaches w'. Thus, no matter how far w is from w', when w circles w' it must always accumulate the same monodromy! In the notation we defined in earlier chapters we have

$$[R^{\text{“}I\text{”}}_{qh-qh}]^2 = e^{2\pi i(\frac{1}{4m}-\frac{1}{8})}$$

Recall that if $a \times b \to c$ we should have $[R^c_{ab}]^2 = e^{2\pi i(h_c-h_a-h_b)}$. Here, the total scaling dimension of the quasihole is $h_{qh} = 1/16 + 1/(8m)$ with the second piece from the bose vertex operator $V_{1/2\sqrt{m}}$. The fusion product “I” $= V_{1/\sqrt{m}}$ has quantum dimension $h_{\text{“}I\text{”}} = 1/2m$.

[14] See work by Bonderson et al

[15] Strictly speaking on the right hand side we should also write the identity operator I for the Ising theory and $V_{1/\sqrt{m}}$ for the boson sector.

Fusion and Braiding of Two Quasiholes in ψ Channel (odd number of electrons)

Let us now assume that the number of electrons is odd. In this case the two σ's of the quasiholes fuse to ψ as in Eq. 17.5. As the two quasiholes approach each other we then have[16]

$$\psi_{qh}(w)\psi_{qh}(w') \sim (w-w')^{\frac{1}{4m}+\frac{3}{8}}$$

where the $\frac{1}{4m}$ is written expliclty in the first line of Eq. 17.9 and the $\frac{3}{8}$ is from the operator product expansion Eq. 17.5. Again we just read off the monodromy from this OPE. Thus, one obtains a different phase depending on the fusion channel of the two quasiholes. In the notation we defined in earlier chapters we have

$$[R^{\text{“}\psi\text{”}}_{qh-qh}]^2 = e^{2\pi i(\frac{1}{4m}+\frac{3}{8})}$$

17.5 Multiple Fusion Channels and Conformal Blocks

We will next address the issue of what happens when we have more than two quasiholes. It is clear what will happen here, we will obtain a correlator (like that in Eq. 17.9) but now it will have more σ fields. We will thus have to figure out how to make sense of correlators with many (nonabelian) σ fields. As an example to show how this works, let us get rid of the ψ fields for a moment and consider a correlator

$$G(w_1,w_2,w_3,w_4) = \langle \sigma(w_1)\sigma(w_2)\sigma(w_3)\sigma(w_4)\rangle \tag{17.10}$$

Let us imagine that we will bring w_1 close to w_2 and w_3 close to w_4. Now in order for the correlator to give a nonzero value, the four fields have to fuse to unity (rule (4) from section 17.2). There are two different ways in which this can happen

$$\sigma(w_1)\sigma(w_2) \to I$$
$$\sigma(w_3)\sigma(w_4) \to I$$

OR we could have

$$\sigma(w_1)\sigma(w_2) \to \psi$$
$$\sigma(w_3)\sigma(w_4) \to \psi$$

and the two ψ fields could then fuse to the identity.

So which one is right? In fact both happen at the same time! To understand this we should think back to what we know about a 2d systems with nonabelian quasiparticles in them — they are described by a vector space. In order to know which particular wavefunction we have in a vector space we need some sort of initial condition or space-time history. Nowhere in the correlator have we specified any space-time history, so we should be getting a vector space rather than a single wavefunction. The multiple wavefunctions in the vector space arise from choosing different roots of the branch cuts of the holomorphic functions. To see a detailed example of this let us write out the explict form of the correlator in Eq. 17.10. We note that the calculation that leads to this requires some substantial knowledge of conformal field theory and will not be presented here. However many of these sorts of results have simply been tabulated in books and can be looked up when necessary. For simplicity we take the four coordinates of the z variables to be at convenient points so that the correlator looks as simple as possible[17].

$$\lim_{w\to\infty}\langle \sigma(0)\sigma(z)\sigma(1)\sigma(w)\rangle = a_+G_+(z) + a_-G_-(z) \tag{17.11}$$

where

$$G_\pm = (wz(1-z))^{-1/8}\sqrt{1\pm\sqrt{1-z}} \tag{17.12}$$

[16] Strictly speaking on the right hand side we should also write the identity operator ψ for the Ising theory and $V_{1/\sqrt{m}}$ for the boson sector.

[17] In fact due to conformal invariance, knowing the correlator for any fixed three points and one point z free, we can determine the correlator for any other four points, but this is beyond the scope of the current discussion!

are known as **conformal blocks** and here a_+ and a_- are *arbitrary* complex coefficients (usually with some normalization condition implied). I.e, the correlator itself represents not a function, but a vector space (with basis vectors being conformal blocks) with arbitrary coefficients yet to be determined by the history of the system!

Let us analyze some limits to see which fusion channels we have here. Taking the limit of $z \to 0$ we find that

$$\lim_{z\to 0} G_+ \sim z^{-1/8} \qquad (\sigma(0)\sigma(z) \to I) \tag{17.13}$$

$$\lim_{z\to 0} G_- \sim z^{3/8} \qquad (\sigma(0)\sigma(z) \to \psi) \tag{17.14}$$

Thus (comparing to Eqs. 17.4 and 17.5) we see that G_+ has $\sigma(0)$ and $\sigma(z)$ fusing to I whereas G_- has them fusing to ψ. Since the four σ's must fuse to the identity, this tells us also the fusion channel for $\sigma(1)$ and $\sigma(w)$.

The most general wavefunction is some linear combination (a_+ and a_-) of the two possible fusion channels. This is what we expect, the state of a system can be any superposition within this degenerate space.

Now consider what happens as we adiabatically take the coordinate z in a circle around the coordinate 1. Looking at Eq. 17.12 we see that we accumulate a phase of $e^{-2\pi i/8}$ from the factor of $(1-z)^{-1/8}$ outside the square-root. In addition, however, the $\sqrt{1-z}$ inside the square root comes back to minus itself when z wraps around 1, thus turning G_+ to G_- and vice versa! The effect of monodromy (taking z around 1) is then

$$\begin{pmatrix} a_+ \\ a_- \end{pmatrix} \longrightarrow e^{-2\pi i/8} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_+ \\ a_- \end{pmatrix}$$

(This result should be somewhat familiar from the homework exercise on Ising anyons!)

We thus see that in this language, the multiple fusion channels are just different choices of which Riemann sheet we are considering, and the fact that braiding (monodromy) changes the fusion channel is simply the fact that moving coordinates around on a Riemann surface, you can move from one Riemann sheet to another!

So long as we can *assume* that the conformal blocks are orthonormal (see comment above on "crucial assumption".. this is now adding a further assumption[18]) then we can continue to read off the result of physically braiding the particles around each other by simply looking at the branch cuts in the wavefunction.

F-matrix

We have seen how to describe the fusion of $\sigma(0)$ and $\sigma(z)$. What if now we instead take z close to 1 such that we can perform an operator product expansion of $\sigma(z)\sigma(1)$. Taking this limit of Eq 17.12 it naively looks like both

$$\lim_{z\to 1} G_+ \sim (1-z)^{-1/8}$$
$$\lim_{z\to 1} G_- \sim (1-z)^{-1/8}$$

But examining this a bit more closely we realize we can construct the linear combinations

$$\tilde{G}_+ = \frac{1}{\sqrt{2}}(G_+ + G_-)$$
$$\tilde{G}_- = \frac{1}{\sqrt{2}}(G_+ - G_-)$$

[18] As with the discussion above, this assumption appears to be true, but "proofs" of it always boil down to some statement about some exotic plasma being in a screening phase, which is hard to prove.

where here we have inserted the prefactor of $1/\sqrt{2}$ such that the new basis $\tilde{G}_\pm$ is orthonormal given that the old basis $G_\pm$ was. With this new basis we now have the limits

$$\begin{aligned}\lim_{z\to 1}\tilde{G}_+ &\sim (1-z)^{-1/8}\\ \lim_{z\to 1}\tilde{G}_- &\sim (1-z)^{-1/8}\left[\sqrt{1+\sqrt{1-z}}-\sqrt{1-\sqrt{1-z}}\right]\\ &\sim (1-z)^{-1/8}(1-z)^{1/2}\sim (1-z)^{3/8}\end{aligned}$$

Thus we see that in this twiddle basis ($\tilde{G}_\pm$) we have in this limit that $\tilde{G}_+$ is the fusion of $\sigma(z)$ and $\sigma(1)$ to identity and $\tilde{G}_-$ is the fusion to ψ.

The transformation between the two bases $G_\pm$ and $\tilde{G}_\pm$ is precisely the F-matrix transformation.

$$\begin{pmatrix}\tilde{G}_+\\ \tilde{G}_-\end{pmatrix}=\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}G_+\\ G_-\end{pmatrix}$$

which should look familiar to anyone who did the homework! (We also got the same result from writing the ising theory in terms of cabled Kauffman strings). Diagrammatically this transform is shown in Fig. 17.1

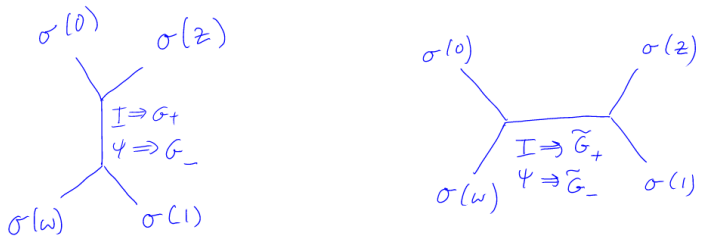


Figure 17.1: The F-matrix transforms between the two fusion channels depicted here.

17.6 More Comments on Moore-Read State with Many Quasiholes

Although we have presented this discussion about multiple fusion channels and braiding in terms of σ operators, the situation is extremely similar once we use quasihole operators ($\sigma(z)V_\beta(z)$) and we put them in a wavefunction as in Eq. 17.9 but possibly with more quasihole operators. As we might expect just from looking at the fusion rules, the number of fusion channels (the number of Riemann sheets!) is $2^{M/2-1}$ where M is the number of quasiholes, and the -1 arises because the overall fusion channel must be the identity. Further, the F-matrices and braiding properties all follow very much in a similar manner. The only slightly problematic piece is that we must continue to assume that the conformal blocks form an orthonormal basis — which is hard to prove, but appears to be true.

17.7 Generalizing to Other CFTs

The principles we used for buidling a quantum Hall state from the Ising CFT can be generalized to build quantum Hall states from other CFTs as well. The general principles are as follows:

(1) Construct an electron field which gives a ground state which is single valued in the electron coordinates. This is done bystarting with an abelian field from the CFT (one that does not have

multiple fusion channels) and combining it with a chiral bose vertex operator. The filling fraction is determined entirely by the charge on the vertex operator.

(2) Identify all of the possible quasiholes by looking at all the fields in the CFT and fusing them with a chiral bose vertex operator and enforcing the condition that the electron coordinates must not have branch cuts. The charge of the quasihole is determined by the charge on the vertex operator (and the charge on the electron vertex operator).

(3) Some of the braiding properties can be determined immediately from the operator product expansion while others require more detailed information about the form of the CFT.

17.7.1 $\mathbb{Z}_3$ Parafermions (briefly)

As an example, let us consider the $\mathbb{Z}_3$ Parafermion CFT. Its primary fields and fusion rules are given by

	h
ψ_1	2/3
ψ_2	2/3
σ_1	1/15
σ_2	1/15
ϵ	2/5

$\times$	ψ_1	ψ_2	σ_1	σ_2	ϵ
ψ_1	ψ_2				
ψ_2	I	ψ_1			
σ_1	ϵ	σ_2	$\sigma_2+\psi_1$		
σ_2	σ_1	ϵ	$I+\epsilon$	$\sigma_1+\psi_2$	
ϵ	σ_2	σ_1	$\sigma_1+\psi_2$	$\sigma_2+\psi_1$	$I+\epsilon$

These fusion rules might look very complicated, but in fact they can be thought of as an abelian $\mathbb{Z}_3$ theory (with fields $I, \psi_1, \psi_2 = \bar{\psi_1}$) fused with a Fibonacci theory (with fields I and τ). We then have

$$\sigma_1 = \psi_2\tau \tag{17.15}$$
$$\sigma_2 = \psi_1\tau \tag{17.16}$$
$$\epsilon = \tau \tag{17.17}$$

and using the Fibonacci fusions $\tau \times \tau = I + \tau$ and the $\mathbb{Z}_3$ fusions $\psi_i \times \psi_j = \psi_{(i+j) \mod 3}$ with ψ_0 being the identity, we recover the full fusion table[19].
Let us propose an electron field

$$\psi_e(z) = \psi_1(z) V_{\sqrt{m+\frac{2}{3}}}(z)$$

where m is a nonnegative integer (even for bosons, odd for fermions). It is easy to check from the OPE that

$$\psi_e(z)\psi_e(w) \sim (z-w)^m \psi_2(z) V_{2\sqrt{m+\frac{2}{3}}}(z)$$

The resulting wavefunction is then

$$\Psi = \langle \psi_e(z_1)\psi_e(z_2)\dots\psi_e(z_N)\, Q\rangle$$

which is known as the Read-Rezayi $\mathbb{Z}_3$ parafermion wavefunction.

The filling fraction of the wavefunction is determined by the vertex operator and is given by

$$\nu = \frac{1}{m+\frac{2}{3}}$$

For the $m = 0$ case this is $\nu = 3/2$ bosons, while for the $m = 1$ case this is $\nu = 3/5$ fermions.

For the case of $m = 0$ it is easy to check that the wavefunction does not vanish when two particles come to the same point, nor does it vanish when three particles come to the same point, but it does vanish when four particles come to the same point. Thus the wavefunction is an exact (densest) zero energy ground state of a *four* particle delta function.

[19] Note that the scaling dimensions h also work out modulo 1. The τ field has $h_\tau = 2/5$ If you add this to $h = 2/3$ for the ψ field you get $h = 2/5 + 2/3 = 1 + 1/15$

While there are 4-particle interactions for these systems for which wavefunctions are the exact ground state, it turns out that there are physically relevant cases where the Read-Rezayi $\mathbb{Z}_3$ parafermion wavefunction is an extremely good trial wavefunction. For bosons interacting with a simple two body δ-function potential potential at filling fraction $\nu = 3/2$,the $\mathbb{Z}_3$ parafermion wavefunction is extremely good. For electrons interacting with simple coulomb interaction (in realistic quantum well samples), it turns out that the wavefunction is extremely good for $\nu = 2 + 2/5$, which we need to particle-hole conjugate in the partly filled Landau level to get a $\nu = 3/5$ wavefunction.

To construct a quasihole we can try building a quasihole from any of the primary field operators. It turns out the one with the lowest charge is constructed from σ_1

$$\psi_{qh}(z) = \sigma_1(z)V_\beta(z)$$

Using the OPE we have

$$\sigma_1(w)\psi_1(z) \sim (z-w)^{-1/3}\epsilon(z)$$

We thus choose

$$\beta = \frac{p}{3\sqrt{m+\frac{2}{3}}}$$

with the smallest charge quasihole then being $p = 1$. With this choice, for a quasihole at position w we generate a factor of

$$\prod_i (z-w)^{1/3}$$

meaning the charge of the quasihole is

$$e^* = e\nu/3$$