Topological Quantum: Lecture Notes and Proto-Book

Steven H. Simon (and Joost K. Slingerland)

CLARENDON PRESS • OXFORD
2017

Some thoughts about this book

This book originated as part of a lecture course given at Oxford in the fall of 2016 and then again in 2017.

The idea of this book is to give a general introduction to topological quantum ideas. This includes topological quantum field theories, topological quantum memories, topological quantum computing, topological matter and topological order — with emphasis given to the examples of toric code, loop gases, string nets, and particularly quantum Hall effects. The book is aimed at a physics audience (i.e., we avoid the language of category theory like the plague!), although some mathematicians may also find the perspectives presented here to be useful.

How to read this book

The book was originally written to be read roughly sequentially. However, you may be able to jump around quite a bit depending on your interests. When the toric code is introduced, it is quite independent of the prior chapters on the general structure of TQFTs.

I should also mention that chapter 3 introduces some basic mathematics that many people may know, and even if you don't, you might be able to carry on without fully reading this chapter (maybe just return to it if you get confused!).

There are often small hitches and caveats that are swept under the rug in the name of simplifying the discussion. I try to footnote these caveats when they occur. Later in the book (in an appendix?) we introduce the bookkeeping rules for TQFTs more carefully. Anyone needing an authoritative set of usable rules for calculations should refer to that part of the book.

A list of useful references with commentary is given at the end of the book.

Contents

Introduction: History of Topology, Knots, Peter Tait and Lord Kelvin

1

The field of topological matter inhabits a beautiful nexus between mathematics, computer science and physics. Within the field of physics, it has been fundamental to a number of subfields. On the one hand, topology and topological matter are key concepts of modern condensed matter physics[1]. Similarly, in the field of quantum information and quantum computation, topological ideas are extremely prominent[2]. At the same time much of our modern study of topological matter is rooted in ideas of topological quantum field theories that developed from the high energy physics, quantum gravity[3], and string theory community starting in the 1980s. These earlier works have even earlier precidents in physics and mathematics. Indeed, the historical roots of topology in physics date all the way back to the 1800s which is where we will begin our story.

In 1867 Lord Kelvin[4] and his close friend Peter Tait were interested in a phenomenon of fluid flow known as a smoke ring[5] — configurations of fluid flow where lines of vorticity form closed loops as shown in Fig. 1.1. Peter Tait built a machine that could produce smoke rings, and showed it to Kelvin who had several simultaneous epiphanies. First, he realized that there should be a theorem (now known as Kelvin's circulation theorem) stating that in a perfectly dissipationless fluid, lines of vorticity are conserved quantities, and the vortex loop configurations should persist for all time. Unfortunately, few dissipationless fluids exist — and the ones we know of now, such as superfluid helium at very low temperatures, were not discovered until the next century[6]. However, at the time, scientists incorrectly believed that the entire universe was filled with a perfect dissipationless fluid, known as Luminiferous Aether, and Kelvin wondered whether one could have vortex loops in the Aether.

At the same time, one of the biggest mysteries in all of science was the discreteness and immutability of the chemical elements. Inspired by Tait's smoke ring demonstration, Kelvin proposed that different atoms corresponded to different knotting configurations of vortex lines in the Aether. This theory of vortex atoms was appealing in that it gave a

[1] The 2016 Nobel Prize was awarded to Kosterlitz, Thouless, and Haldane for the introduction of topological ideas into condensed matter physics. The topic of this book is a great-granddaughter of some of those ideas. In chapter *** we will discuss some of the key works that this Nobel prize honored.

[2] We will see this in chapters *** below.

[3] See chapter 6.

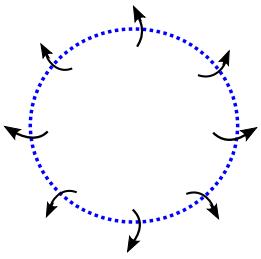


Fig. 1.1 A smoke ring is an invisible ring in space where the fluid flows around the invisible ring as shown by the arrows. The whole thing moves out of the plane of the page at you as the fluid circulates.

[5] A talented smoker can produce a smoke ring from their mouth.

[6] In fact Helium was not even discovered yet in 1867!

[4] Actually, in 1867 he was just William Thomson, but would later be elevated to the peerage and take the name Lord Kelvin after the River Kelvin that flowed by his laboratory.

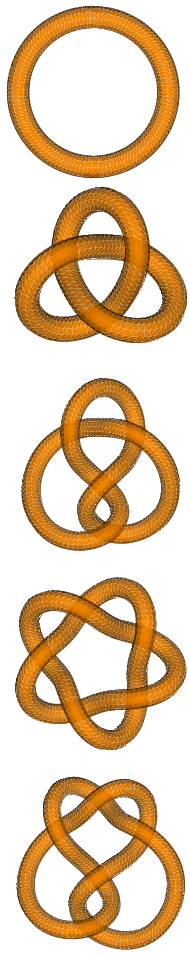


Fig. 1.2 The simplest few knots made from one strand of string. The top knot, a simple loop, is known as the "unknot", and corresponds to the simple smokering in Fig. 1.1. The second knot from the top, known as the trefoil, is not the same as its mirror image (see exercise 2.1)

reason why atoms are discrete and immutable — on the one hand there are only so many different knots that one can make. (See for example, the list of the simplest few knots you can form from on piece of string shown in Fig. 1.2.) On the other hand, by Kelvin's circulation theorem, the knotting of the vortices in a dissipationless fluid (the Aether) should be conserved for all time. Thus, the particular knot could correspond to a particular chemical element, and this element should never change to another one. Hence the atoms should be discrete and immutable!

For several years the vortex theory of the atom was quite popular, attracting the interest of other great scientists such as Maxwell, Kirchhoff, and J. J. Thomson (no relation). However after further research and failed attempts to extract predictions from this theory, the idea of the vortex atom lost popularity.

Although initially quite skeptical of the idea, Tait eventually came to believe that by building a table of all possible knots (knotted configuration of strands such that there are no loose ends) he would gain some insight into the periodic table of the elements, and in a remarkable series of papers he built a catalogue of all knots with up to 7 crossings (the first few entries of the table being shorn in Fig. 1.2). From his studies of knots, Tait is viewed as the father of the mathematical theory of knots, which has been quite a rich field of study since that time (and particularly during the last fifty years).

During his attempt to build his periodic table of knots, Tait posed on what has become perhaps the fundamental question in mathematical knot theory: how do you know if two pictures of knots are topologically identical or topologically different. In other words, can two knots be smoothly deformed into each other without cutting any of the strands. Although this is still considered to be a difficult mathematical problem, a powerful tool that helps answer this question is the idea of a knot invariant which we will study in the next chapter. Shortly, it will become clear how this idea is related to physics.

Although Tait invented a huge amount of mathematics of the theory of knots[7] and developed a very extensive "periodic table of knots", he got no closer to understanding anything about the periodic table of the atoms. In his later life he became quite frustrated with his lack of progress in this direction and he began to realize that understanidng the atoms was probably unrelated to understanding knots. Tait died[8] in 1901 not realizing how much his work on the theory knot would be important in physics, albeit for entirely different reasons.

[7] Some of his conjectures were *way* ahead of their time — some being proven only in the 1980s or later! See A. Stoimenow (Bull. Amer. Math. Soc. 45 (2008), 285–291 or https://arxiv.org/pdf/0704.1941.pdf) for a review of the Tait conjectures proven after 1985.

[8] Peter Tait was also a huge fan of golf and wrote some beautiful papers on the trajectory of golf balls. His son, Freddie Tait, was a champion amateur golfer, being the top amateur finisher in the British Open six times and placing as high as third overall twice. Freddie died very young, at age 30, in the Boer wars in 1900. This tragedy sent Peter into a deep depression from which he never recovered.

Kauffman Knot Invariant and Relation to Physics

2

The purpose of this chapter is to introduce you to a few of the key ideas and get you interested in the subject!

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]. For example, the picture of a knot (or more properly, the picture of the link of two strings) on the left of Fig. 2.1 is topologically equivalent to the picture on the right of Fig. 2.1.

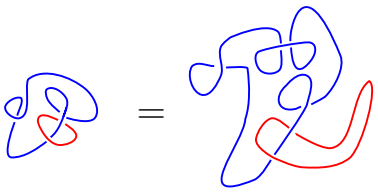


Fig. 2.1 Topological equivalence of two knots. The knot on the left can be deformed continuously into the knot on the right without cutting any strands.

While it appears easy to determine whether two simple knots are topologically equivalent and when they are not, for more complicated knots, it becomes *extremely difficult* to determine equivalence.

A **Knot Invariant** is a mapping from a knot (or a picture of a knot) to an output via a set of rule which are cooked up in such a way that two topologically equivalent knots give the same output. (See Fig. 2.2)

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 this work in very simple terms.[3]

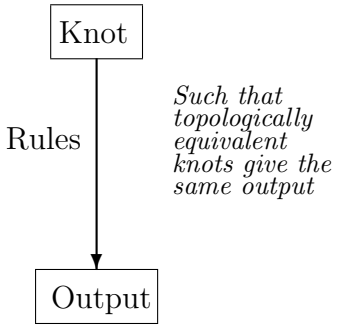


Fig. 2.2 Schematic description of a knot invariant as a set of rules taking an input knot to some mathematical output such that topologically equivalent knots give the same output.

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

[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. A **knot** can be defined as a particular embedding of a circle (S^1) into a three dimensional reference manifold such as $\mathbb{R}^3$ (regular 3-dimensional space) with no self-intersections. A **link** is an embedding of several circles into the three dimensional manifold with no intersections. (2) When I say "topologically equivalent" here I mean the concept of **regular isotopy** (See section 2.2.1 and 3.4) . 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 in section 2.2.1, 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.

[3] Kauffman also wrote a very nice book "Knots and Physics" which I recommend highly.

and replaced with the coefficient

$$d = -A^2 - A^{-2} \tag{2.1}$$

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. This type of replacement rule is known as a *skein* rule.[4] The Kauffman rules are shown in Fig. 2.3.

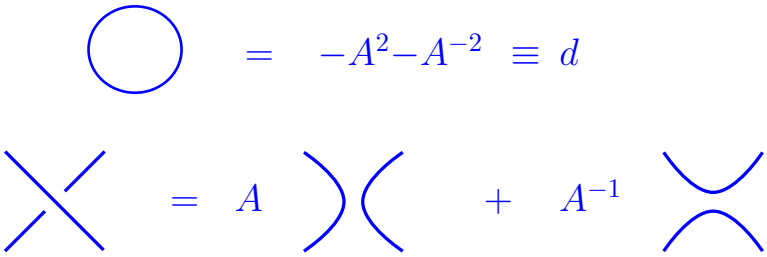


Fig. 2.3 Rules for evaluating the Kauffman Invariant.

The general scheme is to use the second rule of Fig. 2.3 to remove all crossings of a diagram. In so doing, one generates a sum of many diagrams with various coefficients. Then once all crossings are removed, one is just left with simple loops, and each loop can just be replaced by a factor of d.

To give an example of how these rules work we show evaluation of the Kauffman invariant for the simple knot in the upper left of figure 2.4. 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 original 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[5] and still the 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[6]. However, a bit of thought indicates that things still get rapidly difficult for complicated knots. In the example of Fig. 2.4 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.

[4] The word "skein" is an infrequently used English word meaning loosely coiled yarn, or sometimes meaning an element that forms part of a complicated whole (probably both of these are implied for our mathematical usage). "Skein" also means geese in flight, but I suspect that is unrelated.

[5] To a mathematician the Kauffman invariant is an invariant of regular isotopy — see section 2.2.1 below.

[6] The converse is not true. If two knots give the same output, they are not necessarily topologically equivalent. It is an open question whether there are any knots besides the simple unknot which has Kauffman invariant d.

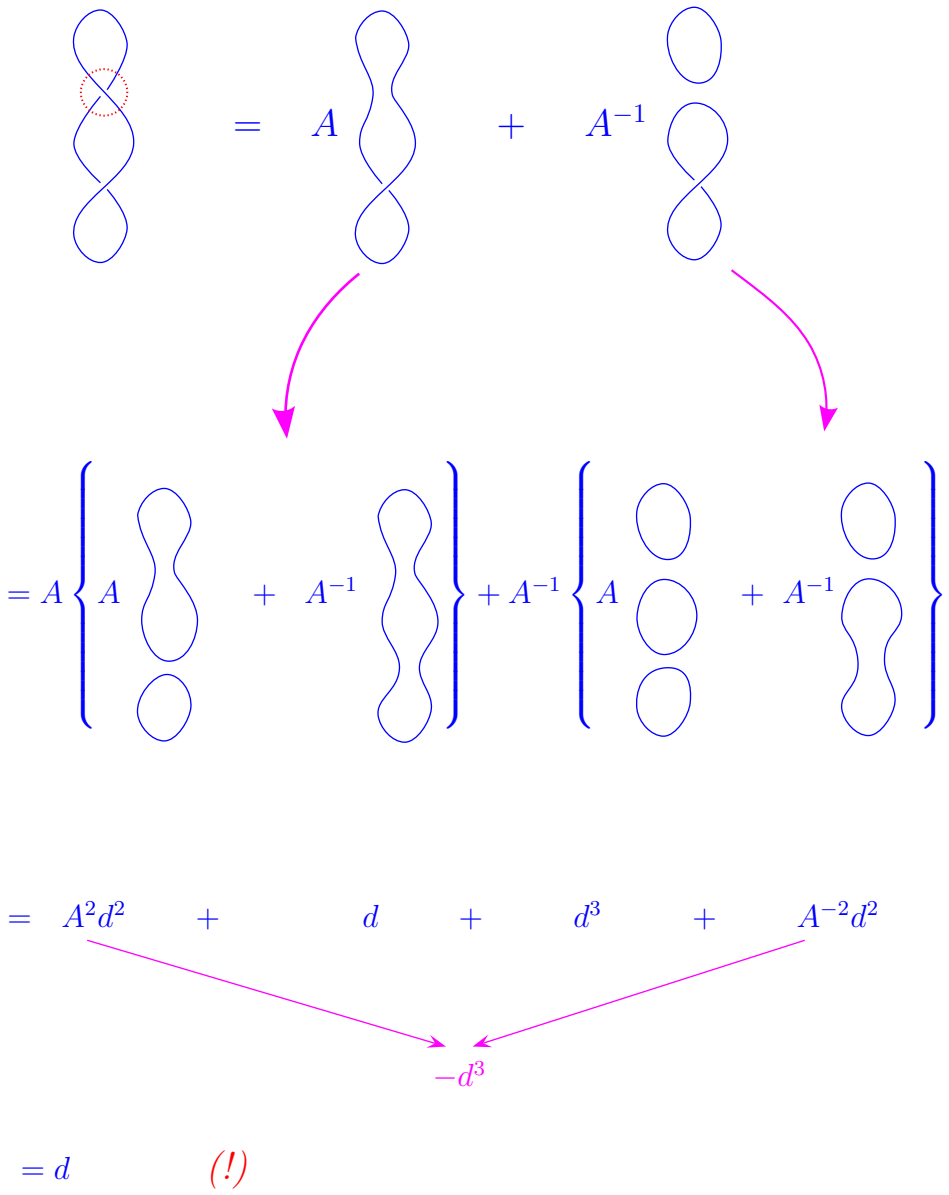


Fig. 2.4 Example of evaluation of the Kauffman invariant for the simple twisted loop in the upper left. The light dotted red circle is meant to draw attention to where we apply the Kauffman crossing rule first to get the two diagrams on the right hand side. After applying the Kauffman rules again, we 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 original knot is just a simple loop (the so-called "unknot") and the Kauffman rules tell us that a loop gets a value d.

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 invariant.

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 (and higher dimensions as well!) but more is known about 2+1 D and we will focus on that at least for now.[7]

[7]There is also some discussion of "topological" systems in 1+1 D later in section ***.

Figure 2.5 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 try to reannihilate. At the end of the process, it is possible that the particles do reannihilate to the vacuum (as shown in the diagram), but it is also possible that (with some probability amplitude) the particle-antiparticle pairs form bound states that do not annihilate back to the vacuum.

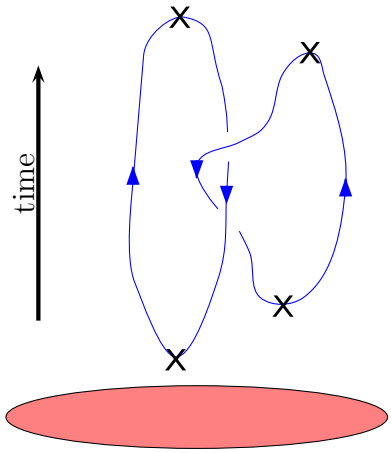


Fig. 2.5 A space-time process showing world lines of particles for a 2+1 dimensional system (shown as the shaded disk at the bottom). The X's mark the points in space-time where particles-anti-particle pairs are either pair-created or pair-annihilated.

In a topological theory, the quantum amplitude for these processes depends on the topology of the world lines, and not on the detailed geometry (I.e., the probability that the particles reannihilate versus form bound states). 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 2.5. It should surprise us that systems exist where amplitudes depend only on topolgy, 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 topology of the space-time paths.

What should be obvious here is that the quantum amplitude of a process is 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[8]. He won the Fields medal along with Vaughn Jones for this work.

[8]See works by Witten, item 6 in Reference List

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 several 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). A brief table of some of the physical systems that are believed to be related to nontrivial knot invariants is given in Table 2.1.

(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 is also known as "Ising" anyons[a]. Possibly physical realizations include

- $\nu = 5/2$ 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[b]
- 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) $SU(2)_1$ **Kauffman class** Also known as semions. These are proposed to be realized in rotating boson fractional quantum Hall effect (although not convincingly produced in experiment yet. See comments in chapter ***). This corresponds to a fairly trivial knot invariant as we will see later in section ***.

(5) $SU(3)_2$ **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.

[a] The Ising conformal field theory, describes the critical point of the 2D classical Ising model. We will discuss the relationship between conformal field theory and topological theories in chapters 19 and ***

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

Table 2.1 Table of some interesting topological systems related to knot invariants. Note that these are closely related to, but not precisely the same as $SU(2)_k$ Chern-Simons theory (which we discuss in chapter refchap:ABCS). NEED TO SYNC This up with the right definitions of the Kauffman category. Cross check with Zhenghan's book.

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 many other quantum hall states that are also topological, but have corresponding knot invariants are fairly trivial, as we will later see in ***.

2.2.1 Twist and Spin-Statistics

Before moving on, let us 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 (as shown in figure 2.6) and try to evaluate its Kauffman invariant.

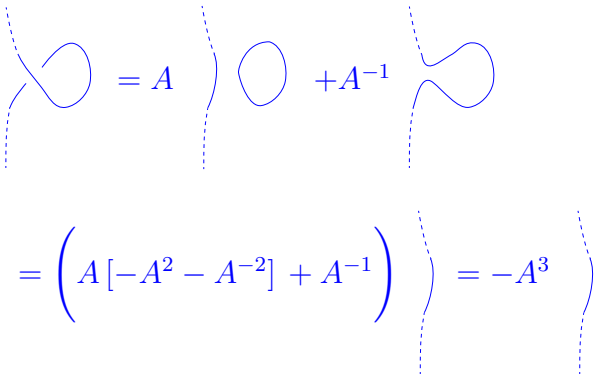


Fig. 2.6 Evaluation of a twist loop in a string. The dotted lines going off the top and bottom of the diagrams mean that the string will be connected up with itself, but we are not concerned with any part of the knot except for piece shown. The result of this calculation is that removal of the little twist in the loop incurs a factor of $-A^3$.

We see from the calculation, that the little loop in the string has value of $-A^3$ compared to a straight string. But this seems to contradict what we said earlier! We 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!

The issue here is that the unlooped string on the right and the looped string on the left are, in fact, *not* topologically equivalent[9]. 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"[10]. If we imagine straightening a thick string (not an infinitely thin string) we realize that pulling it straight gives a twisted string (see fig 2.7) — anyone who has

[9] In mathematics we say they are ambient isotopic but not regular isotopic! (See section 3.4)

[10] We should thus think of our knots as not just being a simple embedding of a circle S^1 into a three manifold $\mathbb{R}^3$, but rather an embedding of a ribbon. This is equivalent to specifying an orthogonal vector at each point along knot which gives the orientation of the ribbon cross section at each point. When one draws an knot as a line, one must have a convention as to what this means for the orientation of the ribbon. See section 3.4.

tried to straighten a garden hose will realize this!

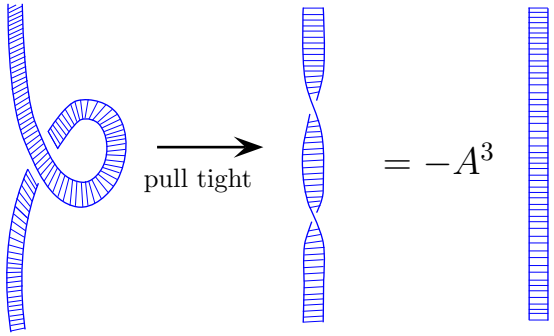


Fig. 2.7 Pulling straight a small loop introduces a twist in the string. This twist can be replaced with a factor of $-A^3$.

So the looped 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 represents 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 2π rotation, and indeed this factor of $-A^3$ is precisely such a phase in any well defined quantum theory.

In fact, figure 2.7 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 chapter 4, 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[11].

As a side comment, one can easily construct a knot invariant that treats the looped string on the left of Fig.2.6 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[12]. This gives the famed Jones Polynomial knot invariant.

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 imagine we tell you everything that you can locally measure about these particles (their positions, if 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 2.8.

[11] In the most interesting case of non-Abelian statistics, there may be multiple possible exchange phases for two particles, although this does not effect the equivalence of diagrams stated here. We will discuss this more in chapter 4.

[12] To properly count the self twists, one calculates the so-called "writhe" of the knot. Give the string an orientation (a direction to walk along the string) and count +1 for each positive crossing and -1 for each negative crossing where a positve crossing is when, traveling in the direction of the string that crosses over, one would have to turn left to switch to the string that crosses under. If we orient the twisted string on the left of Fig. 2.6 as up-going it then has a negative crossing by this definition.

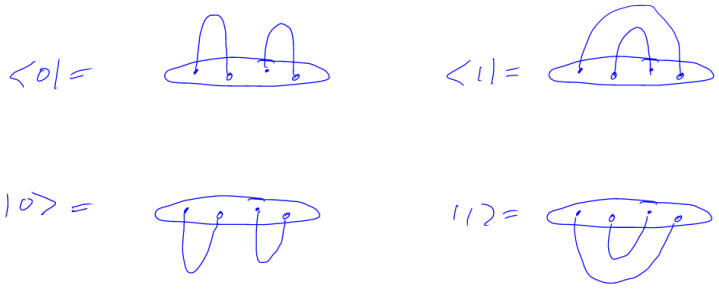


Fig. 2.8 Two linearly independent quantum states that look identical locally but have different space-time history. The horizontal plane is a space-time slice, and the diagrams are all ordiented so time runs vertically. Kets are turned into bras by reversing time.

To demonstrate that these two different space-time histories are linearly independent quantum states, we simply take inner products as shown in Fig. 2.9 by gluing together a ket with a bra. Since $\langle 0|0\rangle = \langle 1|1\rangle = d^2$ but $\langle 0|1\rangle = d$, 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.)

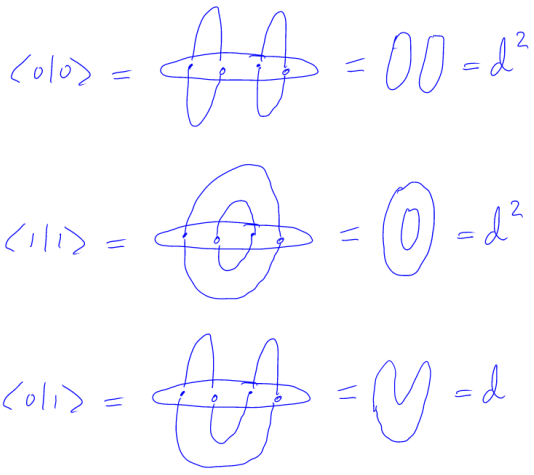


Fig. 2.9 Showing that the kets $|0\rangle$ and $|1\rangle$ are linearly independent. For $|d| \neq 1$ the inner products show they must be linearly independent quantities.

We can even do something more complicated, like insert a braid between the bra and the ket. See Figure 2.10. The braid does a unitary operation on the two dimensional vector space spanned by $|0\rangle$ and $|1\rangle$.

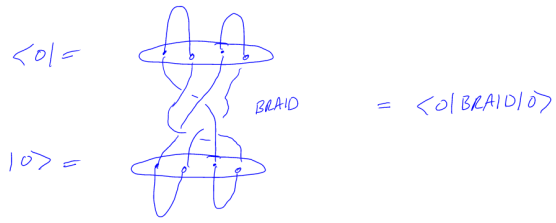


Fig. 2.10 Inserting a braid between the bra and the ket. The braid performs a unitary operation on the two dimensional vector space spanned by $|0\rangle$ and $|1\rangle$

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 (up to the issue of having properly normalized states) the inner product produced by graphical gluing a bra to a ket 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![13] Indeed, quantum computers are really nothing more than qubits 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[14]. The

[13] 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!

[14] 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 calculation 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} dimensional matrix. But if you had the physical system 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

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. 2.11. 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[15]) 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 of a large knot is essentially impossible for a classical computer, for almost all values of A.. However, this is an easy task if you happen to have a topological quantum system in your lab![16] Thus the topological quantum system has computational ability beyond that of a classical computer.

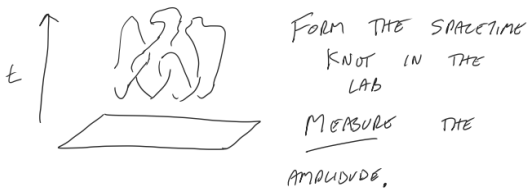


Fig. 2.11 To evaluate the Kauffman invariant: Pull particle-hole pairs from the vacuum, drag them around to form the knot and reannhilate. The probability that they all reannhilate to the vacuum is related to the Kauffman invariant. Thus by repeating the measurement, you can make an estimate of the Kauffman invariant.

It turns out that the ability to calculate Kauffman invariant is sufficient to be able to do any **quantum computation**. One can use this so-called **topological quantum computer** to run algoritms such as Shor's famous factoring (i.e., code breaking) algorithm[17]. The idea of using topological systems for quantum computation is due to Michael Freedman and Alexei Kitaev[18].

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.

[15]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 to the vacuum introduces a $1/\sqrt{d}$ factor in front of the Kauffman invariant.

[16]The details of this are a bit subtle and are discussed in detail by Dorit Aharonov, Vaughan Jones, and Zeph Landau, arXiv:quant-ph/0511096, and also Dorit Aharonov, and Itai Arad arXiv:quant-ph/0605181, New J. Phys. 13 (2011) 035019, and also by Greg Kuperberg, arXiv:0908.0512.

[17]See book by Nielson and Chuang for more detail about quantum computation in general.

[18]Freedman is another Fields 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.

So it turns out that these topological systems can do quantum computation. Why is this a good way to do quantum computation?[13]. 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 in section *** below 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. 2.12. However, as long as the noise does not change the topology of the knot, then no error is introduced. Thus 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.)

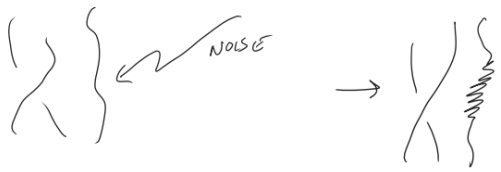


Fig. 2.12 The effect of noise on a topological quantum computation. As long as the noise does not change the topology of the knot, then no error is introduced.

2.5 Some quick comments about Fractional Quantum Hall Effect

There will be chapters later about Fractional Quantum Hall Effect (FQHE). But it is worth saying a few words about FQHE as a topological system now.

FQHE occurs in two dimensional electronic systems[19] 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 in chapter ***). The FQHE state emerges at low temperature and is topological[20].

[19] 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

[20] 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

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.2.13. 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.

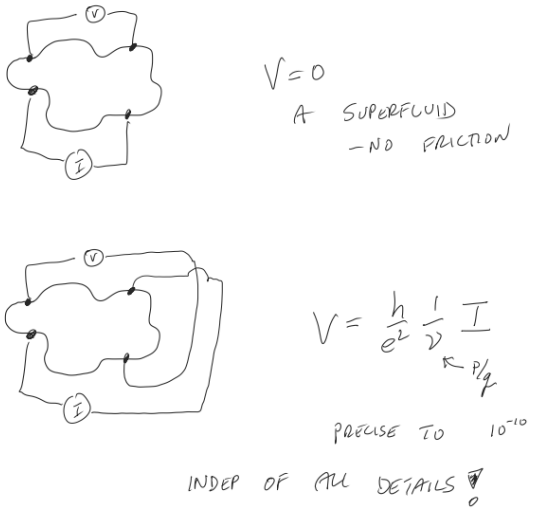


Fig. 2.13 Measurement of resistance in FQHE experiment.

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 of V/I 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 (soon to be) 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 mea-

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.

sure the resistance of a bar of copper, the voltage would depend entirely on how far apart you 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.

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 Summary

- Knot invariants, such as the Kauffman invariant, help distinguish knots from each other.
- The quantum dynamics of certain particles are determined by certain knot invariants.
- Computation of certain knot invariants is computationally "hard" on a classical computer, but not hard using particles whose dynamics is given by knot invariants.
- Computation by braiding these particles is equivalent to any other quantum computer.
- Physical systems which have these particles include fractional quantum Hall effect.

Further Reading

Louis Kauffman, *Knots and Physics*, World Scientific, (2001), 3ed.

Steven H. Simon,"Quantum Computing with a Twist", *Physics World*, September 2010, p35-40.

Exercises

Exercise 2.1 Trefoil Knot and the Kauffman (Jones) Invariant Using the Kauffman rules, calculate the Kauffman Invariant of the Right and Left handed Trefoil knots. Conclude these two knots are topologically inequivalent. While this statement appears obvious on sight, it was not proved mathematically until the 1900s. (It is trivial using this technique!).

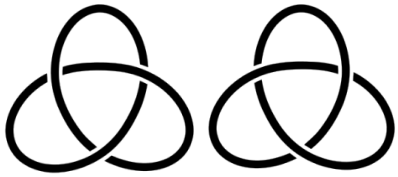
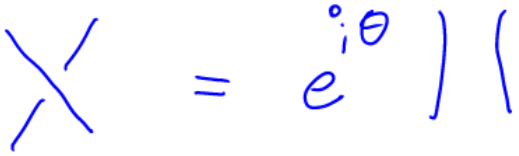


Fig. 2.14 Left and Right Handed Trefoil Knots (on the left and right respectively)

Exercise 2.2 Abelian Kauffman Anyons

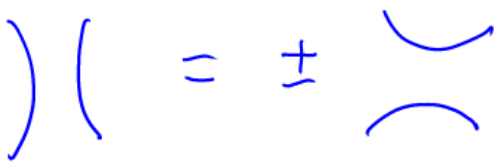
Anyons described by the Kauffman invariant with certain special values of the constant A are abelian anyons – meaning that an exchange introduces only a simple phase.



(a) For $A = \pm e^{i\pi/3}$ (and the complex conjugates of these values), show that the anyons are bosons or fermions respectively (i.e., $e^{i\theta} = \pm 1$).

(b) For $A = \pm e^{i\pi/6}$ (and the complex conjugates of these values) show the anyons are semions (i.e., $e^{i\theta} = \pm i$). In fact these are precisely the anyons that arise for the $\nu = 1/2$ fractional quantum Hall effect of bosons (We will discuss this later in the term. While it is still controversial whether this particular phase of quantum Hall matter has been produced experimentally as of now, it is almost certain that it will be produced experimentally and convincingly within a few years.)

HINT: Aim to show first in the two respective cases that



If you can't figure it out, try evaluating the Kauffman invariant for a few knots with these values of A and see how the result arises.

Some Mathematical Basics

3

It is a shame that the Oxford undergrads do not get any proper education in advanced mathematics. To make up for this failing, I am including a very short exposition of most of what you need to know. For much of the book, you won't even need to know this much! To be honest, I recommend skipping this chapter and referring back to the relevent sections only when you need it!

The section on manifolds is used mostly in sections ***

The section on groups is used in sections ***

3.1 Manifolds

We sometimes write $\mathbb{R}$ to denote the real line, i.e., it is a space where a point is indexed by a real number x. We can write $\mathbb{R}^n$ to denote n-dimensional (real) space — a space where a point is indexed by n-real numbers $(x_1, \ldots, x_n)$. Sometimes people call these spaces "Euclidean" space.

Definition 3.1 *A* **Manifold** *is a space that locally looks like a Euclidean space.*

3.1.1 Some Simple Examples: Euclidean Spaces and Spheres

- $\mathbb{R}^n$ is obviously a manifold.
- The circle S^1, also known as a 1-sphere (hence the notation, the index 1 meaning it is a 1-dimensional object) is defined as as all points in a plane equidistant from a central point. Locally this looks like a line since position is indexed by a single variable (the "curvature" of the circle is not important locally). Globally, one discovers that the circle is not the same as a real line, as position is periodic (if you walk far enough in one direction you come back to where you start). We sometimes define a circle as a real number from 0 to 2π which specifies the angle around the circle.
- The 2-sphere S^2 is what we usually call (the surface of) a sphere in our regular life. We can define this similarly as all points in $\mathbb{R}^3$ equidistant from a central point.
- One can generally define S^n as points equidistant from a central point in $\mathbb{R}^{n+1}$.

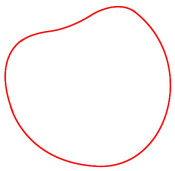


Fig. 3.1 This object is topologically a circle, S^1.

Often when we discuss a manifold, we will often be interested in its topological properties only. In other words, we will not care if a circle is dented as shown in Fig. 3.1, it is still topologically S^1. Mathematicians say that two objects that can be smoothly deformed into each other are *homeomorphic*, although we will not use this language often.

It is sometimes convenient to view the circle S^1 as being just the real line $\mathbb{R}^1$ with a single point added "at infinity" — think about joining up $+\infty$ with $-\infty$ to make a circle. We can do the same thing with the sphere S^2 and $\mathbb{R}^2$ — this is like taking a big flat sheet and pulling the boundary together to a point to make it into a bag (which is a sphere S^2). Obviously the idea generalizes, S^3 is the same as $\mathbb{R}^3$ "compactified" with a point at infinity, and so forth.

3.1.2 Unions of Manifolds $\mathcal{M}_1 \cup \mathcal{M}_2$

We can take a "disjoint" union of manifolds. For example, $S^1 \cup S^1$ is two circles (not connected in any way). If we think of this as being a single manifold, it is a manifold made of two disjoint pieces. Locally it still looks like a Euclidean space.

3.1.3 Products of Manifolds: $\mathcal{M}_3 = \mathcal{M}_1 \times \mathcal{M}_2$

One can also "cross" two manifolds together $\mathcal{M}_3 = \mathcal{M}_1 \times \mathcal{M}_2$. This means that a point in $\mathcal{M}_3$ is given by one point in $\mathcal{M}_1$ and one point in $\mathcal{M}_2$. This multiplication is often called the *direct* or *Cartesian* product.

- $\mathbb{R}^2 = \mathbb{R}^1 \times \mathbb{R}^1$. Here, a point in $\mathbb{R}^1$ is specified by a single real number. Crossing two of these together, a point in $\mathbb{R}^2$ is specified by two real numbers (one in the first $\mathbb{R}^1$ and one in the second $\mathbb{R}^1$.)
- $T^2 = S^1 \times S^1$. The two torus T^2, or surface of a dougnut (donut, if you are from the states and you like coffee) is the product of two circles. To see this note that a point on a torus is specified by two angles, and the torus is periodic in both directions. Similarly we can build higher dimensional tori (tori is the plural of torus) by crossing S^1's together any number of times.

3.1.4 Manifolds with Boundary:

One can also have manifolds with boundary. A boundary of a manifold locally looks like an n-dimensional half-Euclidean space. The interior of a manifold with boundary looks like a Euclidean space, and near the boundary it looks like a half-space, or space with boundary . For example, a half-plane is a 2-manifold with boundary. Probably an example is useful:

- The n-dimensional ball, denoted B^n is defined as the set of points in n dimensional space such that the distance to a central point is less than or equal to[1] some fixed radius r. Note: Often the

[1]This defines an "closed" ball. The "open" ball would be the set of points less than the radius. Physicists are not often concerned with this difference, but to be precise, we will usually talk about closed manifolds.

Ball is called a disk and is denoted by D^n (so $D^n = B^n$). The nomenclature makes good sense in two dimensions, where what we usually call a disc is D^2. The one-dimensional ball is just an interval (one-dimensional segment) which is sometimes denoted I.

Note that a boundary of a manifold may have disconnected parts. For example, the boundary of an interval (segment) in 1-dimension $I = B^1$ is two disconnected points at its two ends. $\partial I = \text{pt} \cup \text{pt}$ where pt means a point and here $\cup$ means the union of the two objects as described above in 3.1.2.

One can take cartesian products of manifolds with boundaries too. For example, consider the interval (or 1-ball) $I = B^1$ which we can think of as all the points on a line with $|x| \leq 1$. The cartesian product $I \times I$ is described by two coordinates (x, y) where $|x| \leq 1$ and $|y| \leq 1$. This is a square including its interior. However, in topology we are only ever concerned with topological properties, and a square-with-interior can be continuously deformed into a circle-with-interior, or a 2-ball (2-disc), B^2.

- The same reasoning gives us the general topological law $B^n \times B^m = B^{n+m}$.
- The cylinder (hollow tube) is expressed as $S^1 \times I$ (two coordinates, one periodic, one bounded on both sides).
- The solid donut is expressed as $B^2 \times S^1$, a (2)-disc crossed with a circle.

3.1.5 Boundaries of Manifolds: $\mathcal{M}_1 = \partial\mathcal{M}_2$.

The notation for boundary is ∂, so if $\mathcal{M}_1$ is the boundary of $\mathcal{M}_2$ we write $\mathcal{M}_1 = \partial\mathcal{M}_2$. The boundary ∂M has dimension one less than that of $\mathcal{M}$.

- The boundary of D^2 the 2 dimensional disc is the circle S^1.
- More generally, the boundary of B^n (also written as D^n) is S^{n-1}.

It is an interesting topological principle that the boundary of a manifold is always a manifold without boundary. Or equivalently, the boundary of a boundary is the empty set. We sometimes write $\partial^2 = 0$ or $\partial(\partial\mathcal{M}) = \emptyset$ where $\emptyset$ means the empty set.

- The boundary of the 3-dimensional ball B^3 is the sphere S^2. The sphere S^2 is a 2-manifold without boundary.

The operation of taking a boundary obeys the Leibnitz rule analogous to taking derivatives

$$\partial(\mathcal{M}_1 \times \mathcal{M}_2) = (\partial\mathcal{M}_1) \times \mathcal{M}_2 \quad \cup \quad \mathcal{M}_1 \times (\partial\mathcal{M}_2)$$

Lets see some examples of this:

- Consider the cylinder $S^1 \times I$. Using the above formula with find its boundary

$$\partial(S^1 \times I) = (\partial S^1) \times I \ \cup \ S^1 \times \partial I = S^1 \cup S^1$$

To see how we get the final result here, start by examining the first term. Here, S^1 has no boundary so $\partial S^1 = \emptyset$ and therefore everything before the $\cup$ symbol is just the empty set. In the second term the boundary of the interval is just two points $\partial I = \text{pt} \cup \text{pt}$. Thus the second term gives the final result $S^1 \cup S^1$, the union of two circles.

- Consider writing the disc (topologically) as the product of two intervals $B^2 = I \times I$. It is best to think of this cartesian product as forming a filled-in square. Using the above formula we get

$$\begin{aligned} \partial B^2 &= \partial(I \times I) = (\text{pt} \cup \text{pt}) \times I \ \cup \ I \times (\text{pt} \cup \text{pt}) \\ &= (I \cup I) \cup (I \cup I) = \text{top} \cup \text{bottom} \cup \text{left} \cup \text{right} \\ &= \text{square (edges only)} = S^1 \end{aligned}$$

The formula gives the union of four segments denoting the edges of the square.

3.2 Groups

A **group** G is a set of elements $g \in G$ along with an operation that we think of as multiplication. The set 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 writing $g_1 g_2$ 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 **Abelian**[2]. 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**[3]

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[4] 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

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

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

[3] Apparently named after someone named NonAbel.

[4] 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. For more abstract groups, e is often most natural.

3.2.1 Some Examples of Groups

- The group of integers $\mathbb{Z}$ with the operation being addition. The identity element is 0. This group is Abelian.
- 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. This group is Abelian.
- The group $\mathbb{Z}_N$ which is the set of complex numbers $e^{2\pi ip/N}$ with p an integer (which can be chosen between 1 and N inclusive) and the operation being multiplication. This is equivalent to the set of integers modulo N with the operation being addition. This group is Abelian.
- The group of permutations of N elements, which we write as S_N (known as the **permutation group**, or **symmetric group**). This group is Non-Abelian. There are $N!$ elements in the group. Think of the elements of the group as being a one-to-one mapping from 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. Here we choose a convention that we do the operation written furthest right first. You can choose either convention, but then you must stick to it! You will see both orderings in the literature!) So, if we start with the element 1, when we 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

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

Note that if we multiply the elements in the opposite order we get a different result (hence this group is Non-Abelian)

$$g_4 = g_1 g_2 = \begin{cases} 1 & \to & 1 \\ 2 & \to & 3 \\ 3 & \to & 2 \end{cases}$$

3.2.2 Lie Groups and Lie Algebras

A **Lie group**[5] is a group which is also a manifold. Roughly, a group with a continuous (rather than discrete) set of elements. Examples include:

- The group of invertible $n \times n$ 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 invertible $n \times n$ *real* matrices. We call this group $GL(n, \mathbb{R})$.
- The group, $SU(2)$, the set of 2 by 2 unitary matrices with unit determinant. In this case the fact that this is also a manifold can be made particularly obvious. We can write all $SU(2)$ matrices as

$$\begin{pmatrix} x_1 + ix_2 & -x_3 + ix_4 \\ x_3 + ix_4 & x_1 - ix_2 \end{pmatrix}$$

  with all x_j any real numbers with the constraint that $x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1$. Obviously the set of four coordinates (x_1, x_2, x_3, x_4) with the unit magnitude constraint describes the manifold S^3.
- $SU(N)$, the group of unitary N by N matrices of determinant one is a Lie group
- $SO(N)$, the group of real rotation matrices in N dimensions is a Lie group.
- The vector space $\mathbb{R}^n$ with the operation being addition of vectors, is a Lie group.

Note that certain Lie groups are known as "simple" because as manifolds they have no boundaries and no nontrivial limit points (For example, $GL(n)$ is not simple because there is a nontrivial limit — you can continuously approach matrices which have determinant zero (or are not invertable) and are therefore not part of the group. The set of simple Lie groups (including, $SU(N)$ and $SO(N)$ and just a few others) is extremely highly studied.

A **Lie Algebra** is the algebra generated by elements infinitesimally close to the identity in a Lie group[6]. For matrix valued Lie groups G, we can write any element $g \in G$ as

$$g = e^X = \mathbf{1} + X + (X)^2/2 + \dots$$

where X is an element of the corresponding Lie algebra (make it have small amplitude such that g is infinitesimally close to the identity). Conventionally if a Lie group is denoted as G the corresponding Lie algebra is denoted $\mathfrak{g}$.

- For the Lie group $SU(2)$, we know that a general element can be written as $g = \exp(i\mathbf{n}\cdot\boldsymbol{\sigma})$ where $\mathbf{n}$ is a real three-dimensional vector and $\boldsymbol{\sigma}$ are the Pauli matrices. In this case $i\sigma_x$, $i\sigma_y$ and $i\sigma_z$ are the three generators of the Lie algebra $\mathfrak{su}(2)$ (in the fundamental representation).

[5] Pronounced "Lee", named after Sophus Lie, also a Norwegian Mathematician of the 1800s. Like Ski-Jumping, Norway seems to punch above its weight in the theory of groups.

[6] A slightly more rigorous definition is that a Lie algebra is an algebra of elements $u, v, w...$ which can be added with coefficients a, b, c to give $X = au + bv + cw + \dots$ where we have a commutator $[\cdot, \cdot]$ which satisfies $[X, X] = 0$ for all X as well as bilinearity $[au + bv, X] = a[u, X] + b[v, X]$ and similarly $[X, au + bv] = a[X, u] + b[X, v]$ for all X, a, b, u, v, and finally we must have the Jacobi identity $[[X, Y], Z] + [[Y, Z], X] + [[Z, X], Y] = 0$.

- For the Lie group $GL(n,\mathbb{R})$ the corresponding Lie algebra $\mathfrak{gl}(n,\mathbb{R})$ is just the algebra of $n \times n$ real matrices.

3.2.3 Representations of Groups:

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 $g_1, g_2 \in G$.

Typically in quantum mechanics we are concerned with representations which are unitary, i.e., $\rho(g)$ is a complex unitary matrix of some dimension. (In case you don't remember, a unitary matrix U has the property that $UU^\dagger = U^\dagger U = \mathbf{1}$).

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

3.3 Fundamental Group $\Pi_1(\mathcal{M})$

A powerful tool of topologists is the idea of the fundamental group of a manifold $\mathcal{M}$ which is often called the first homotopy group, or $\Pi_1(\mathcal{M})$. This is essentially the group of topologically different paths through the manifold starting and ending at the same point.

First, we choose a point in the manfiold. Then we consider a path through the manifold that starts and ends at the same point. Any other path that can be continuously deformed into this path (without changing the starting point or ending point) is deemed to be topologically equivalent (or homeomorphic, or in the same equivalence class). We only want to keep one representative of each class of topologically distinct paths.

These topologically distinct paths form a group. As one might expect, the inverse of a path (always starting and ending at the same point) is given by following the same path in a backward direction. Multiplication of two paths is achieved by following one path and then following the other to make a longer path.

3.3.1 Examples of Fundamental Groups

- If the manifold is a circle S^1 the topologically distinct paths (starting and ending at the same point) can be described by the number n of clockwise wrappings the path makes around the circle before

coming back to its starting point (note n can be 0 or negative as well). Thus the elements of the fundamental group are indexed by a single integer. We write $\Pi_1(S^1) = \mathbb{Z}$.

- If the manifold is a torus S^1 the topologically distinct paths can be described by two integers indicating the number of times the path winds around each handle. We write $\Pi_1(S^1 \times S^1) = \mathbb{Z} \times \mathbb{Z}$.

It is in fact, easy to prove that $\Pi_1(\mathcal{M}_1 \times \mathcal{M}_2) = \Pi_1(\mathcal{M}_1) \times \Pi_1(\mathcal{M}_2)$.

- A fact known to most physicists is that the the group of rotations of three dimensional space $SO(3)$ is not simply connected — a 2π rotation (which seems trivial) cannot be continuously deformed to the trivial rotation, whereas a 4π rotation can be continuously deformed to the trivial rotation.[7] Correspondingly the fundamamental group is the group with two elements $\Pi_1(SO(3)) = \mathbb{Z}_2$.

[7] This is the origin of half-odd integer angular momenta.

3.4 Isotopy, Reidermeister Moves, and Blackboard Framing

We ran into the idea of isotopy in chapter 2. Two knots (or two pictures of knots) are isotopic if one can be deformed into each other other without cutting any of the strands. Usually this is referred to as "ambient isotopy". In order for two pictures of knots to be ambient isotopic they must be related to each other by a series of moves, known as Reidermeister moves[8].

[8] This is a very old result of knot theory: Reidemeister, Kurt (1927), "Elementare Begrndung der Knotentheorie", Abh. Math. Sem. Univ. Hamburg, 5 (1): 2432. Note that it may take many many moves in order to bring a knot into some particular desired form. For example, if there are c crossings in a diagram which is equivalent to the simple unknot (an unknotted loop), the strongest theorem yet proven is that it can be reduced to the simple unknot with $(236c)^{11}$ moves. (Lackenby, Marc (2015), "A polynomial upper bound on Reidemeister moves", Annals of Mathematics, Second Series, 182 (2): 491564, MR 3418524, arXiv:1302.0180.

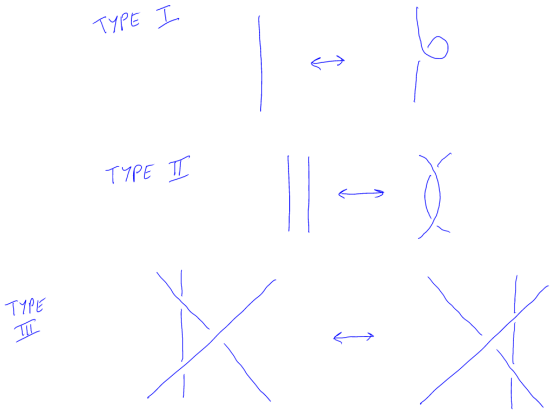


Fig. 3.2 The Three Reidermeister Moves. Any two knots that can be deformed into each other without cutting (they are "ambient isotopic) can be connected by a series of Reidermeis¿ter moves.

It is also useful to define *regular isotopy* which is when two knots can be connected using only type-II and type-III moves. Another way of thinking about this is to think of the strings as being ribbons. A

type-I move inserts a twist in the ribbon (See Fig. 2.7) and gives back a different ribbon diagram, whereas type-II and III moves do not twist the ribbon.

Since it is important to specify when a strand of string has a self-twist (as in the middle of Fig. 2.7) it is a useful convention to use so-called *blackboard framing*. With this convention we always imagine that the string really represents a ribbon and the ribbon always lies in the plane of the blackboard. An example of this is shown in Fig. 3.3.

Fig. 3.3 Blackboard framing. The knot drawn on the left represents the ribbon on the right, where the ribbon always lies flat in the plane of the blackboard.

Chapter summary

Some mathematical ideas introduced in this chapter:

- **Manifolds** are locally like Euclidean space: Examples include sphere S^2, circle S^1, torus surface $T^2 = S^1 \times S^1$, etc. Manifolds can also have boundaries, like a two dimensional disk B^2 (or D^2) bounded by a circle.
- **Groups** are mathematical sets with an operation, and identity and an inverse: Important examples include, $\mathbb{Z}$ the integers under addition, $\mathbb{Z}_N$ the integers mod N under addition, the symmetric (or permutation group) on N elements S_N, and Lie groups such as $SU(2)$ which are also manifolds at the same time as being groups.
- **The Fundamental Group** of a manifold is the group of topologically different paths through the manifold starting and ending at the same point.
- **Isotopy** is the topological equivalence of knot diagrams (what can be deformed to what without cutting

Further Reading

For background on more advanced mathematics used by physicists, including some topological ideas, see:

- M. Nakahara, *Geometry, Topology, and Physics, 2ed*, (2003), Taylor and Francis.

- M. Stone and P. Goldbart, *Mathematics for Physics*, Cambridge (2009). Free pdf prepublication version available online.

For further information on mathematics of knots, isotopy, and Riedermeister moves, see

- Louis Kauffman, *Knots and Physics*, World Scientific, (2001), 3ed.

Particle Quantum Statistics

4

In chapter 2 we discussed braiding particles around each other, 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 two bosons the wavefunction is unchanged, if we exchange two fermions the wavefunction accumulates a minus 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 the identity, or one. There are two square roots of one: +1 and -1, so these are the only two possibilities for the exchange operator.

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, it is very useful to think about quantum physics from the Feynman path integral point of view.[3]

4.1 Single Particle Path Integral. (Probably you know this!)

Consider a space-time trajectory of a non-relativistic particle. 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.

[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] Although based on ideas by Pauli, 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.

[3] If you are familiar with path integrals you can certainly skip down to section 4.2. If you are not familiar with path integrals, please do not expect this to be a thorough introduction! What is given here is a minimal introduction to give us what we need to know for our purposes and nothing more!

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

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

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 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 infinitessimally small pieces, and you end up doing lots of integrals over all possible intermediate positions. In order to get the final result, you must sum over all 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

$$\langle \mathbf{x}_f | \hat{U}(t_f, t_i) | \mathbf{x}_i \rangle = \mathcal{N} \sum_{\substack{\text{paths } \mathbf{x}(t) \text{ from} \\ (\mathbf{x_i}, t_i) \text{ to } (\mathbf{x_f}, t_f)}} e^{iS[\mathbf{x}(t)]/\hbar} \tag{4.1}$$

where $\mathcal{N}$ is some normalization constant. Here $S[\mathbf{x}(t)]$ is the (classical!) 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. 4.1 is often well defined as a limit of dividing the path into discrete time steps and integrating over $\mathbf{x}$ at each time. We often rewrite this sum over paths figuratively as a so-called path integral

$$\langle \mathbf{x}_f | \hat{U}(t_f, t_i) | \mathbf{x}_i \rangle = \mathcal{N} \int_{(\mathbf{x_i}, t_i)}^{(\mathbf{x_f}, t_f)} \mathcal{D}\mathbf{x}(t) \ e^{iS[\mathbf{x}(t)]/\hbar} \tag{4.2}$$

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 minimum 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 $S[\mathbf{x}(t)]$ is extremized — the function $\mathbf{x}(t)$ which extremizes the action. This is just the classical principle of least action!

4.2 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 $\mathbf{x}_2$. 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 both 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[4]. For simplicity, we can assume that $\mathbf{x}_1 \neq \mathbf{x}_2$, i.e., the particles have hard cores and cannot overlap[5].

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 $|\mathbf{x}_2, \mathbf{x}_1\rangle$ mean physically different things. To construct a path integral, we want to think about all possible paths through this configuration space.

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 (See the discussion in section 3.3).

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. 4.1. 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.

We will call the (left fig) non-exchange path TYPE 1, and the (right fig) exchange path TYPE -1. 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

[4] This ordering scheme works in one dimension. In two dimensions 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.

[5] It is sometimes even more convenient to declare $|\mathbf{x}_1 - \mathbf{x}_2| > \epsilon$.

Fig. 4.1 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)

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.

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

$$= \mathcal{N}\left(\sum_{\substack{\text{TYPE 1 paths}\\ i\to f}} e^{iS[\text{path}]/\hbar} + \sum_{\substack{\text{TYPE -1 paths}\\ i\to f}} e^{iS[\text{path}]/\hbar}\right)$$

This second line is simply a rewriting 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

$$\langle \mathbf{x}_{1f}\mathbf{x}_{2f}|\hat{U}(t_f,t_i)|\mathbf{x}_{1i}\mathbf{x}_{2i}\rangle = \tag{4.3}$$
$$\mathcal{N}\left(\sum_{\substack{\text{TYPE 1 paths}\\ i\to f}} e^{iS[\text{path}]/\hbar} - \sum_{\substack{\text{TYPE -1 paths}\\ i\to f}} e^{iS[\text{path}]/\hbar}\right)$$

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 to see if we still have the composition law obeyed. Again, we break the time propagation at some intermediate time

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

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

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 minus 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. 4.3. 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.

4.3 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 said a bit about the permutation group in the mathemtaical section (**??**) 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 doesn't matter.. We can think of this as N directed curves moving in $ND+1$ dimensional space[6]. Since we want to add up all of these possible paths in a path integral it is useful to try to better understand what the structure is of this space of paths.

[6]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.

Again, 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 (See section 3.3). The reason this is 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.

4.3.1 Paths in 2+1 D, the Braid Group

A path through this configuration space is known as a braid. An example of a braid is shown in Fig.4.2.

Fig. 4.2 A path through configuration space for 3 Particles in 2+1 D is a braid with three strands.

A few notes about braids:

(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[7] on N strands is typically notated as B_N. The generators of the braid group on 4 strands are shown in Fig. 4.3 Any braid

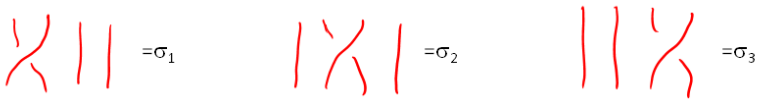


Fig. 4.3 The three generating elements of the braid group on 4 strands, B_4. Any braid on four strands (any element of B_4) can be written as a product of the braid generators and their inverses (the inverse of an element looks similir but the crossing is reversed – counter-clockwise instead of clockwise).

can be written as a product of the braid generators 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. 4.4. A very useful braid

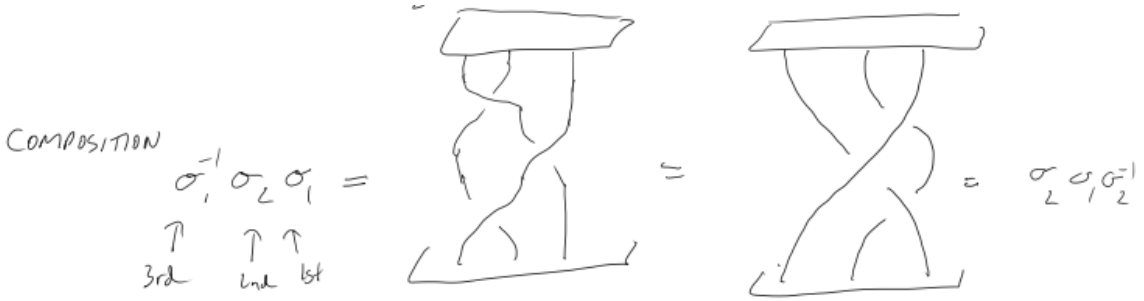


Fig. 4.4 Two braid words that represent the same braid. The observant reader will see the similarity here to Reidermeister moves of type-III discussed above in section 3.4. Similarly $\sigma_i\sigma_i^{-1} = 1$ is a type-II move.

invariant is given by the so-called winding number

$$\begin{aligned} W &= \text{Winding Number} \\ &= \#\text{ of overcrossings} - \#\text{ of undercrossings} \end{aligned}$$

where an overcrossing is a σ and an undercrossing is a σ^{-1}. As can be checked in Fig.4.4, the winding 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.

[7] Sometimes it is also convenient to define the "pure" braid group PB_N which requires that each strand begin and end at the same position. So, for example, σ_1^2 is an element of the pure braid group, but σ_1 is not.

4.3.2 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 of a one-dimensional world-line embedded in a four-dimensional space. If this is not obvious consider the following lower dimensional analogue shown in Fig.4.5.

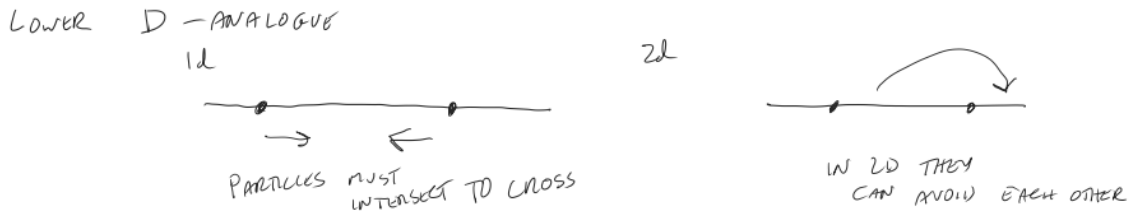


Fig. 4.5 In one dimension, two points cannot cross through each other without hitting each other. But if we allow the points to move in 2D they can get around each other without touching. This is supposed to show you that one-dimensional world-lines cannot form knots in 4d space.

As shown in the figure, in one dimension, 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 fourth dimension, we can move one string a bit off into the fourth 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 three-dimensions 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. 4.6. The only thing that is important is who starts where and ends where. This is precisely the permutation group, or symmetric group S_N (see section **??**). Note that in the symmetric group and exchange squared does give the identity. However, in the braid group this is not so — the braid σ_i^2 is not the identity since it creates a nontrivial braid!

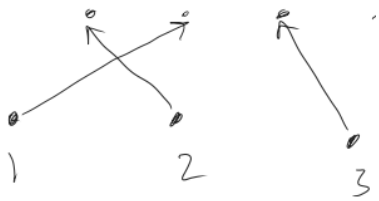


Fig. 4.6 Paths in 3+1 D are elements of the permutation group (or symmetric group) S_N.

4.3.3 Building a Path Integral

We now return to the issue of building a path integral. We will follow the intuition we gained in the two 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

$$\langle\{\mathbf{x}\}_f|\hat{U}(t_f,t_i)|\{\mathbf{x}\}_i\rangle = \mathcal{N}\sum_{g\in G}\rho(g)\sum_{\substack{\text{paths}\in g\\ i\to f}} e^{iS[\text{path}]/\hbar} \tag{4.4}$$

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 *unitary representation* of the group G. (See section 3.2.3 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.

$$\begin{aligned}
&\langle\{\mathbf{x}\}_f|\hat{U}(t_f,t_i)|\{\mathbf{x}\}_i\rangle = \\
&= \int d\{\mathbf{x}\}_m\ \langle\{\mathbf{x}\}_f|\hat{U}(t_f,t_m)|\{\mathbf{x}\}_m\rangle\ \langle\{\mathbf{x}\}_m|\hat{U}(t_m,t_i)|\{\mathbf{x}\}_i\rangle \\
&\sim \int d\{\mathbf{x}\}_m \left(\sum_{g_1\in G}\rho(g_1)\sum_{\substack{\text{paths}\in g_1\\ m\to f}}\right)\left(\sum_{g_2\in G}\rho(g_2)\sum_{\substack{\text{paths}\in g_2\\ i\to m}}\right) e^{iS[\text{path}]/\hbar}
\end{aligned}$$

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_1g_2. The prefactors of the paths $\rho(g_1)$ and $\rho(g_2)$ then multiply and we get $\rho(g_1)\rho(g_2) = \rho(g_1g_2)$ 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 the path. In other words, the form shown in Eq. 4.4 is properly preserved under composition, which is what is required in quantum mechanics!

4.4 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.

4.4.1 3+1 Dimensions

The group of paths through configuration space G is the symmetric group S_N. There are *only two possible* one-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 two particle example we did above where the exchange was assigned a -1.

4.4.2 2+1 Dimensions

The group of paths through configuration space G is the braid group B_N. We can describe the possible one-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 $e^{i\theta}$ whereas a counterclockwise exchange accumulates a phase of $e^{-i\theta}$.

- 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 $e^{i\pi} = e^{-i\pi}$). This is **fermions.**
- **Any** other value of θ is also allowed. This is known as **Anyons**, or **fractional statistics**. They are also known as **Abelian Anyons** in contrast with the Non-Abelian case which we will discuss in a moment.

The fact that this fractional statistics is consistent in quantum mechanics was first point out by Leinaas and Myrheim in 1977[8], and popularized by Wilczek[9] in the early 1980s. In 1984 and 1985 Halperin and then Arovas, Schrieffer, and Wilczek showed theoretically that anyons really occur in fractional quantum Hall systems.

4.5 Non-Abelian Case

Can we do something more interesting and exotic by using 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 $\{\mathbf{x}\}$. Even once we fix the positions (as well as the values of any local quantum

[8] There is no reason why this might not have been discovered in the 1930s, but no one bothered to think about it. It is a bit incredible that experimental systems that realize anyons were discovered so soon after the theoretical proposal, since the original theoretical work was entirely abstract, and they were not thinking about any particular experiment.

[9] Among other things, Wilczek coined the term *anyon.* (He also won a Nobel prize for asymptotic freedom, but that is mostly unrelated.)

numbers, like any "color" or "flavor" degree of freedom associated with the particle), suppose there still remains an M-fold degeneracy of the state of the system. We might describe the M states as $|n;\{\mathbf{x}\}\rangle$ for $n = 1$ to M. An arbitary wavefunction of the system can then be expressed as

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

with the A_n's being some complex coefficients.[10] Given the N positions $\{\mathbf{x}\}$, a general wavefunction should be thought of as a vector in M dimensional complex 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

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

where U is an M by M matrix, and is assumed to 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 (which is a vector). Thus, if an initial state is given as a vector A_n, when we braid particles around each other in some braid g, the vector gets transformed by multiplying it by the matrix $U(g)$ to give the final state vector.

A particle that obeys this type of braiding statistics is known as a **non-Abelian anyon**, or **nonabelion**.[11] The word "non-abelian" means non-commutative, and the term is used since generically matrices (in this case the U matrices) don't commute.

In general the Hilbert space dimension M is exponentially large in the number of particles N. We define a quantity d, know as the **quantum dimension** such that

$$M \sim d^N \tag{4.5}$$

where the $\sim$ means that it scales this way in the limit of large N. We will see a lot more of this quantity d later. It is not coincidence that we used the symbol d previously in the context of Kauffman anyons! (See Eq. 2.1.)

Some Quick Comments on Quantum Computing:

Quantum Computing is nothing more than the controlled application of unitary operations to a Hilbert space[12]. Unitary operations is exaclty what we can do by braiding nonabelions around each other! I.e., we are multiplying a vector by a unitary matrix. Thus we see how braiding of particles, as discussed in chapter 2 can implement quantum computation.[13]

4.5.1 Parastatistics in 3D

Is it possible to have exotic statistics in 3+1D? Indeed, there do exist higher dimensional representations 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

[10] If we want $|\psi\rangle$ normalized then there is a normalization condition on the A_n coefficients. For example, if the $|n\rangle$'s are orthonormal then we need $\sum_n |A_n|^2 = 1$ in order that $|\psi\rangle$ is normalized.

[11] 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 effect.

[12] And initialization and measurement

[13] The observant reader will notice that for quantum computation we are no longer summing over all possible braids, but we are specifying a particular braid that the particles should take in order to implement a particular unitary operation. To do this we must control the paths of the particles, by say, holding them in traps that we move. In principle all paths are still included in the path integral, but only the ones we specify contribute significantly.

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 key 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 on earth 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 how this happens 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. (We return to this in section ***)

We should note that despite this important result, 3+1D is certainly not boring — but in order to get "interesting" examples, we have to relax some of our constraints. For example, if we relax the condition that "particles" are pointlike, but consider string-like objects instead, then we can have exotic statistics that describe what happens when one loop of string moves through another (or when a point-like particle moves through a loop of string).

Chapter Summary

- The path integral formulation of quantum mechanics requires us to add up all possible paths in space time.
- We can add all of these paths in any way that preserves the composition law and the different possibilities allow for different types of particle statistics.
- The topologically different paths of N particles in space-time form a group structure (the fundamental group of the configuration space) which is the permutation group S_N in 3+1 dimensions, but is the braid group B_N in 2+1 dimensions.
- Particle braiding statistics must be a representation of this group.
- In 3+1 dimensions we can only have bosons and fermions, but in 2+1 dimensions we can have nontrivial braiding statistics which may be abelian (or "fractional") or non-abelian.
- Quantum computation can be performed by braiding with certain non-abelian representations.

Further Reading

For information on particle statistics see:

- F. Wilczek, ed. *Fractional Statistics and Anyon Superconductivity*, World Scientific, (1990).
- Chetan Nayak, Steven H. Simon, Ady Stern, Michael Freedman, Sankar Das Sarma, *Non Abelian Anyons and Topological Quantum Computation*, Rev. Mod. Phys. 80, 1083 (2008). Also available online at https://arxiv.org/abs/0707.1889

For a basic primer on path integrals see

- F. Essler, *Lecture Notes for the C6 theory option*, http://www-thphys.physics.ox.ac.uk/people/FabianEssler/C6web2012/lecturenotes2015.pdf
- R. MacKenzie, *Path Integral Methods and Applications*, https://arxiv.org/abs/quant-ph/0004090

Exercises

Exercise 4.1 About the Braid Group (a) Convince yourself geometrically that the defining relations of the braid group are:

$$\sigma_i\,\sigma_{i+1}\,\sigma_i = \sigma_{i+1}\,\sigma_i\,\sigma_{i+1} \qquad 1 \le i \le M-2 \tag{4.6}$$

$$\sigma_i\,\sigma_j = \sigma_j\,\sigma_i \quad \text{for} \quad |i-j|>1, \qquad 1 \le i,j \le M-1 \tag{4.7}$$

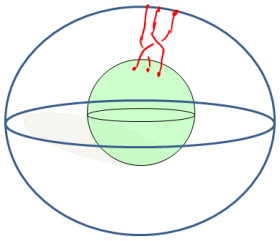


Fig. 4.7 An element of the braid group $B_3(S^2)$. The braid shown here is $\sigma_1\sigma_2^{-1}$

(b) Instead of thinking about particles on a plane, let us think about particles on the surface of a sphere. In this case, the braid group of M strands on the sphere is written as $B_M(S^2)$. To think about braids on a sphere, it is useful to think of time as being the radial direction of the sphere, so that braids are drawn as in Fig. 4.7. The braid generators on the sphere still obey Eqns. 4.6 and 4.7, but they also obey one additional identity

$$\sigma_1\sigma_2\ldots\sigma_{M-2}\sigma_{M-1}\sigma_{M-1}\sigma_{M-2}\ldots\sigma_2\sigma_1 = I \tag{4.8}$$

where I is the identity (or trivial) braid. What does this additional identity mean geometrically?

[In fact, for understanding the properties of anyons on a sphere, Eq. 4.8 is not quite enough. We will try to figure out below why this is so by using Ising Anyons as an example.]

Exercise 4.2 Ising Anyons and Majorana Fermions

The most commonly discussed type of non-Abelian anyon is the Ising anyon (we will discuss this in more depth later). Ising anyons occurs in the Moore-Read quantum Hall state ($\nu = 5/2$), as well as in any chiral p-wave superconductor and in recently experimentally relevant so called "Majorana" systems.

The non-Abelian statistics of these anyons may be described in terms of Majorana fermions by attaching a Majorana operator to each anyon. The Hamiltonian for these Majoranas is zero – they are completely noninteracting.

In case you haven't seen them before, Majorana Fermions γ_j satisfy the anticommutation relation

$$\{\gamma_i, \gamma_j\} \equiv \gamma_i\gamma_j + \gamma_j\gamma_i = 2\delta_{ij} \tag{4.9}$$

as well as being self conjugate $\gamma_i^\dagger = \gamma_i$.

(a) Show that the ground state degeneracy of a system with $2N$ Majoranas is 2^N if the Hamiltonian is zero. Thus conclude that each *pair* of Ising anyons is a two-state system. Hint: Construct a regular (Dirac) fermion operator from two Majorana fermion operators. For example,

$$c^\dagger = \frac{1}{2}(\gamma_1 + i\gamma_2)$$

will then satisfy the usual fermion anti-commutation $\{c, c^\dagger\} = cc^\dagger + c^\dagger c = 1$. (If you haven't run into fermion creation operators yet, you might want to read up on this first!) There is more discussion of this transformation in a later problem on "Ising F-matrix".

(b) When anyon i is exchanged clockwise with anyon j, the unitary transformation that occurs on the ground state is

$$U_{ij} = \frac{e^{i\alpha}}{\sqrt{2}}[1 + \gamma_i\gamma_j] \qquad i < j. \tag{4.10}$$

for some real value of α. Show that these unitary operators form a representation of the braid group. (Refer back to the previous problem, "About the Braid Group"). In other words we must show that replacing σ_i with $U_{i,i+1}$ in Eqns. 4.6 and 4.7 yields equalities. This representation is 2^N dimensional since the ground state degeneracy is 2^N.

(c) Consider the operator

$$\gamma^{\text{FIVE}} = (i)^N\gamma_1\gamma_2\ldots\gamma_{2N} \tag{4.11}$$

(the notation FIVE is in analogy with the γ^5 of the Dirac gamma matrices). Show that the eigenvalues of γ^{FIVE} are ± 1. Further show that this eigenvalue remains unchanged under any braid operation. Conclude that we actually have two 2^{N-1} dimensional representations of the braid group. We will assume that any particular system of Ising anyons is in one of these two representations.

(d) Thus, 4 Ising anyons on a sphere comprise a single 2-state system, or a qubit. Show that by only braiding these four Ising anyons one cannot obtain all possible unitary operation on this qubit. Indeed, braiding Ising anyons is not sufficient to build a quantum computer. [Part (d) is not required to solve parts (e) and (f)]

(e) [bit harder] Now consider $2N$ Ising anyons on a sphere (See above problem "About the braid group" for information about the braid group on a sphere). Show that in order for either one of the 2^{N-1} dimensional representations of the braid group to satisfy the sphere relation, Eqn. 4.8, one must choose the right abelian phase α in Eq. 4.10. Determine this phase.

(f) [a bit harder] The value you just determined is not quite right. It should look a bit unnatural as the abelian phase associated with a braid depends on the number of anyons in the system. Go back to Eqn. 4.8 and insert an additional abelian phase on the right hand side which will make the final result of part (e) independent of the number of anyons in the system. In fact, there should be such an additional factor — to figure out where it comes from, go back and look again at the geometric "proof" of Eqn. 4.8. Note that the proof involves a self-twist of one of the anyon world lines. The additional phase you added is associated with one particle twisting around itself. The relation between self-rotation of a single particle and exchange of two particles is a generalized spin-statistics theorem.

Exercise 4.3 Small Numbers of Anyons on a Sphere

On the plane, the braid group of two particles is an infinite group (The free group with one generator). However, this is not true on a sphere

First review the problem "About the Braid Group" about braiding on a sphere.

(a) Now consider the case of two particles on a sphere. Determine the full structure of the braid group. Show it is a well known finite discrete group. What group is it?

(b) [Harder] Now consider three particles on a sphere. Determine the full structure of the braid group. Show that it is a finite discrete group. [Even Harder] What group is it? It is "well known" only to people who know a lot of group theory. But you can google to find information about it on the web with some work. It may be useful to list all the subgroups of the group and the multiplication table of the group elements.

(c) Suppose we have two (or three) anyons on a sphere. Suppose the ground state is two-fold degenerate. If the braid group is discrete, conclude that no possible type of anyon statistics will allow us to do arbitrary $SU(2)$ rotations on this degenerate ground state by braiding.

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

5

This chapter introduces a simple model of how fractional statistics anyons can arise. After reviewing Aharanov-Bohm effect, we describe these exotic particles as charge-flux composites and explore some of their properties. Finally we see how this fits into the framework of abelian Chern-Simons theory and briefly discuss its non-abelian generalization.

5.1 Review of Aharanov-Bohm Effect

Let us consider the two slit interference experiment shown in Fig. 5.1 We all know the result of the two slit experiment but let us rewrite the calculation roughly as a path integral. We can write

$$\sum_{\text{paths}} e^{iS/\hbar} = \sum_{\text{paths, slit 1}} e^{iS/\hbar} + \sum_{\text{paths, slit 2}} e^{iS/\hbar}$$

$$\sim e^{ikL_1} + e^{ikL_2}$$

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, and k is the wavevector of the incoming wave. In other words, we get the usual two slit calculation.

Now let us change the experiment to that shown in Fig. 5.2. 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.

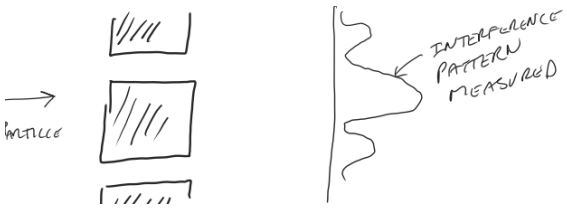


Fig. 5.1 The Young two slit experiment

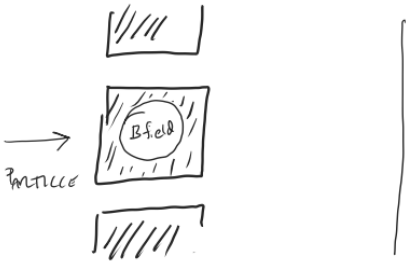


Fig. 5.2 Adding a magnetic field into the hidden box in the Young two slit experiment. Here the circular region includes a constant magnetic field. No magnetic field leaks out of the box. Nonetheless, if the particle being sent into the interferometer is charged, the interference pattern is changed compared 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

$$L = \frac{1}{2m}\dot{\mathbf{x}}^2 + q(\mathbf{A}(\mathbf{x}) \cdot \dot{\mathbf{x}} - A_0) \tag{5.1}$$

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 force as we should expect in electromagnetism.[2] So adding a magnetic field to the Lagrangian can be rephrased as changing the action

$$S \to S_0 + q\int dt\ \dot{\mathbf{x}} \cdot \mathbf{A} = S_0 + q\int \mathbf{dl} \cdot \mathbf{A} \tag{5.2}$$

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

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

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

$$\exp\left[\frac{iq}{\hbar}\int_{\text{slit 1}} \mathbf{dl}\cdot\mathbf{A} - \frac{iq}{\hbar}\int_{\text{slit 2}} \mathbf{dl}\cdot\mathbf{A}\right] = \exp\left[\frac{iq}{\hbar}\oint \mathbf{dl}\cdot\mathbf{A}\right]$$

where the integral is around a loop that goes forward through slit 1 and then backwards through slit 2.

[2]Here are the steps:

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

Using Stokes' theorem, we have

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

where Φ_{enclosed} is the flux enclosed in the loop. Thus there is a measurable relative phase shift between the two paths given by $\frac{iq}{\hbar}\Phi_{\text{enclosed}}$. This results in a shift of the interference pattern measured 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 2π 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. (This is sometimes known as the Aharonov-Casher effect.)

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

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

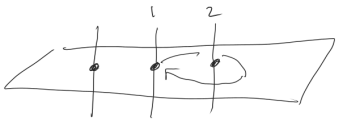


Fig. 5.3 Abelian anyons represented as charges bound to flux tubes through the plane. The charge of each particle is q, the flux of each tube is Φ.

5.2 Anyons as Flux-Charge Composites

We will now consider a simple model of Abelian anyons as charge-flux composites. Imagine we have a two dimensional system with charges q in them, where each charge is bound to an infinitely thin flux tube through the plane, with each tube having flux Φ as shown in Fig. 5.3. If we drag one particle around 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. 5.4)

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

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, giving us anyons as disucssed in chapter 4.

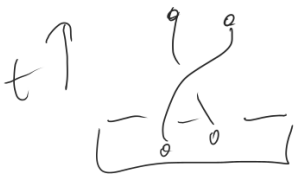


Fig. 5.4 An exchange

Spin of an anyon

Let us see if we can determine the spin of these anyons. Spin refers to properties of 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

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

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. 2.7).

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

5.2.1 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)-(anti-anyon) 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 — negative flux multiplies negative charge). However, the phase of dragging an anyon clockwise around an antianyon is -2θ.

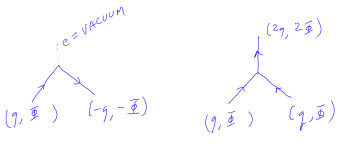


Fig. 5.5 Left; Fusing an anyon and an antianyon to get the vacuum ("e") drawn as dotted line. Note that the antianyon moving forward in time is drawn as a downpointing arrow — which looks like an anyon moving backwards in time. Right: Fusing two anyons together to form another anyon of a different type (two $\theta = q\Phi/\hbar$ anyons make a 4θ anyon)

5.2.2 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. 5.5).

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. 5.5 to indicate that two θ particles have come together to form a 4θ particle.

The principle of locality is an important one. Consider Fig. 5.6 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.

5.3 Anyon Vacuum on a Torus 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 torus (the surface of 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. 5.7. The two cycles around the torus are marked as C_1 and C_2.

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 until 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 until 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 let us consider the operator $T_2^{-1}T_1^{-1}T_2T_1$ where we 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. 5.8.

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 must have unit modulus (i.e., they are just a complex phase). Considering 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 will 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

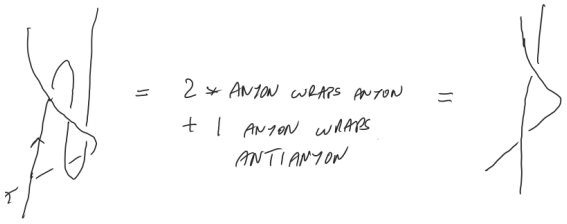


Fig. 5.6 Braiding around a worldline that doubles back. The anyon-antianyon braiding rules are constructed so that the double-back can be pulled straight. The calculation scribbled in the middle says "two anyon wraps around anyon + 1 anyon wraps anti-anyon" which give a net phase the same as one anyon wrapping one anyon.

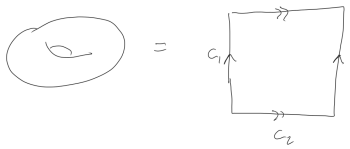


Fig. 5.7 Drawing a torus as a rectangle with opposite edges identified. The two noncontractable cycles around the torus can be considered to be the edges of the square, labeled C_1 and C_2 here.

[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, which will be more complicated.

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

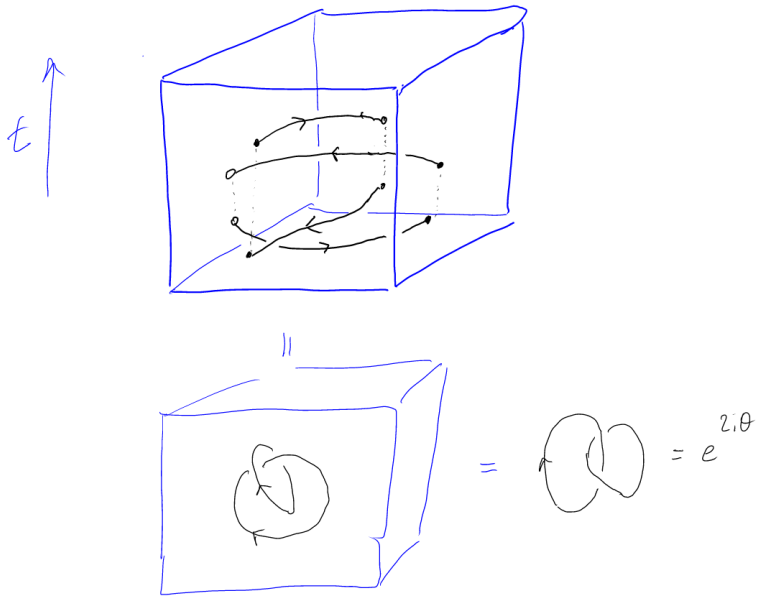


Fig. 5.8 The torus is drawn as a horizontal rectangle with opposite ends identified. Time runs vertically. First create a particle hole pair at the center of the rectangle and move them in opposite directions until they meet at the edges of the rectangle to reannhiliate. Note that a particle moving to 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.

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

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

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

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

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

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

[7] There could be even more degeneracy which would be 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 the anyons in the system. Since we are considering anyons of statistical angle $\theta = \pi p/m$ we can describe this with a charge-flux composite $(q, \Phi) = (\pi p/m, 1)$. Fusion of n of these elementary anyons will have[8]

$$\begin{aligned} \text{Fusion of } n \text{ elementary anyons} &= |\text{“}n\text{”}\rangle = (q = n\pi p/m, \Phi = n) \\ &= (n\pi p/m, n) \end{aligned}$$

[8] By this time I'm sick of writin $\hbar$ and I'm going to set it equal to 1.

Something special happens when we have a cluster of m of these elementary anyons:

$$|\text{“}m\text{”}\rangle = (\pi p, m)$$

If we braid an arbitrary cluster $|\text{“}n\text{”}\rangle = (n\pi p/m, n)$ around one of these $|\text{“}m\text{”}\rangle = (\pi p, m)$ clusters, we obtain a net phase[9] 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 all particles get trivial phase if they braid all the way around $|\text{“}m\text{”}\rangle$.

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

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 an exception. If both p and m are odd, one obtains a nontrivial sign for exchanging (half braiding) a $|\text{“}m\text{”}\rangle = (\pi p, m)$ with another $|\text{“}m\text{”}\rangle = (\pi p, m)$ 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 $|\text{“}m\text{”}\rangle$ 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[10].

[10] Later on we will call this kind of theory “non-modular.” See section ***

Neglecting this more 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 species of particles.

5.3.1 Quantum Memory and Higher Genus

The degenerate ground state on the torus can be thought of as a quantum memory. If there are m different ground states, 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) complex 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”.

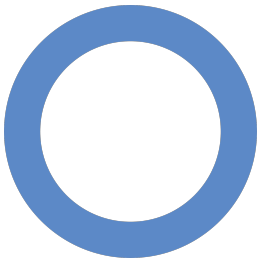


Fig. 5.9 An annulus.

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 as shown in Fig. 5.9). In this case, T_1 could correspond to moving an anyon around loop of the 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 more 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.

5.4 Abelian Chern-Simons Theory

It is useful to see how charge-flux binding occurs in a microscopic field theory description of a physical system. The type of field theory we will study is known as a Chern-Simons field theory[11] and is the paradigm for topological quantum field theories.

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, or simple fractional statistics particles). We start by imagine a gauge field a_α, known as the Chern-Simons vector potential, 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 put 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

$$\Delta\mathcal{L} = \frac{\mu}{2}\epsilon^{\alpha\beta\gamma}a_\alpha\partial_\beta a_\gamma - qj^\alpha a_\alpha \tag{5.3}$$

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

[11] S. S. Chern was one of the most important mathematicians of the 20th century. Jim Simons was 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 world.

values $0, 1, 2$ where 0 indicates the time direction and 1, 2 are the space directions (and j^0 is the particle density).

The first term in Eq. 5.3 is the Lagrangian density of the Chern-Simons vector potential itself. (It is sometimes known as the "Chern-Simons Term"). The second term in Eq. 5.3 couples the Chern-Simons vector potential to the particles in the system. Its form, $j^\alpha a_\alpha$, is actually something we have already seen. If we have N particles then the current is

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

The j^0 component, the density, is just a delta function peak at the position of each particle. The 1 and 2 component, $\mathbf{j}$ 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

$$\sum_{n=1}^{N} q\left[\mathbf{a}(\mathbf{x}_n) \cdot \mathbf{x}_n - a_0(\mathbf{x}_n)\right] \tag{5.4}$$

exactly as in Eq. 5.1. So this is nothing more than the regular coupling of a system of charged particles to a vector potential.

As is usual for a gauge theory, 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 (first term in Eq. 5.3) is gauge invariant on a closed manifold if we can integrate by parts. To see this, make an arbitrary gauge transformation

$$a_\mu \to a_\mu + \partial_\mu \chi \tag{5.5}$$

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 vanishes by antisymmtry. Note that this this gauge invariance 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 in section *** but for now, let us just think about closed space-time manifolds).

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

$$\frac{\partial \mathcal{L}}{\partial a_\alpha} = \partial_\beta \left(\frac{\partial \mathcal{L}}{\partial(\partial_\beta a_\alpha)} \right) \tag{5.6}$$

which generates the equations of motion[12]

$$qj^\alpha = \mu \epsilon^{\alpha\beta\gamma} \partial_\beta a_\gamma \tag{5.7}$$

[12] 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 Eq. 5.3. However, note that when we differentiate with respect to a_α on the left hand side of Eq. 5.6, we also generate an identical factor of $\mu/2$ and these two add up.

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

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

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 Eq. 5.7 by a_α and then plug it back into the Lagrangian 5.3. 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 exactly 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

$$\sum_{\text{paths }\{\mathbf{x}(t)\}}\ \sum_{\text{all }a_\mu(\mathbf{x},t)} e^{i(S_0+S_{CS}+S_{coupling})/\hbar} \tag{5.8}$$

Here the first sum is the usual sum over particle paths that we have discussed before. The second sum is the sum over all possible configurations of the field $a_\mu(\mathbf{x},t)$. 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

$$\sum_{\text{all }a_\mu(\mathbf{x},t)} \rightarrow \int \mathcal{D}a_\mu(x)$$

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

Formally when we write a functional integral we mean[13] that we should divide space and time into little boxes and within each box integrate over all possible values of a_μ. Fortunately, we will not need to do this procedure explicitly.

We thus rewrite Eq. 5.8 as

$$\sum_{\text{paths }\{\mathbf{x}(t)\}} e^{iS_0/\hbar}\int\mathcal{D}a_\mu(x)\ e^{iS_{CS}/\hbar}\ e^{i(q/\hbar)\int_{paths}dl^\alpha a_\alpha} \tag{5.9}$$

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 (from the first term in Eq. 5.3). The final exponential in Eq. 5.9 represents the coupling (from the second term of Eq. 5.3) 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 (see also Eq. 5.4). 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 in section 5.5.

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

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

where W is the winding number of the path and θ is the anyon statistical angle. In other words, integrating out the Chern-Simons gauge field implements fractional statistics for the particles in the system, inserting a phase $e^{\pm i\theta}$ for each exchange!

Something we have pointed out above in section 5.3 is that the vacuum of an anyon theory 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[14]. In fact this integral is exactly the ground state degeneracy of the system if we are considering a three dimensional manifold of the form $\mathcal{M} = \Sigma \times S^1$ for a 2D manifold Σ and compactified time[15]. 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, particularly section 5.6.3). This quantity $Z(\mathcal{M})$, often known as the partition function of the theory for the manifold $\mathcal{M}$, will be of crucial importance as we learn more about topological theories in general in Chapter 7 below.

[14] Some space time manifolds we might consider, such as any 2D manifold Σ cross time (such that $\mathcal{M} = \Sigma \times \mathbb{R}$), seem very natural. However, as we will see in much detail in chapter 7, we will want to be much more general about the types of manifolds we consider. We should even allow three dimensional manifolds where the two-dimensional topology of a fixed time slice changes as time evolves! See also the discussion in 6 and Fig. 6.1.

[15] Compactification of time from $\mathbb{R}$ to S^1 is something that might be familiar from statistical physics where this procedure is used for representing finite temperatures.

5.5 NonAbelian Chern-Simons theory: The paradigm of TQFT

Among 2+1 dimensional topological quantum systems, 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 as we have been doing so far.

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.[16] More precisely to construct a non-Abelian Chern-Simons theory, we consider a vector potential that take values in a Lie algebra[17]. For example, if we choose to work with the Lie

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

[17] See the introduction to Lie groups and Lie algebras in section 3.2.2. In brief: A Lie Group is a group which is also a manifold. A Lie *algebra* is the algebra of infinitesimal changes in this group.

algebra of $SU(2)$ in the fundamental representation we can write a general element of this algebra as a sum of the three generators $i\sigma_x, i\sigma_y, i\sigma_z$ so that our Lie algebra value gauge field is then

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

where σ_k are the Pauli matrices (the factors of 2 and i are a useful convention, although other conventions exist). 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[18] loop operators

$$W_L = \mathrm{Tr}\left[P \exp\left(i\frac{q}{\hbar}\oint_L dl^\mu a_\mu\right)\right] \tag{5.10}$$

where here the integral follows some path L. (The trace can be taken in any representation of the group giving a physically different meaning. See footnote 21 below.) 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, we have a problem that $a_\mu(x)$ 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)

$$P\exp(i\oint_L dl^\mu a_\mu) = \tag{5.11}$$
$$[1 + ia_\mu(x_1)dl^\mu(x_1)]\,[1 + ia_\mu(x_2)dl^\mu(x_2)]\,[1 + ia_\mu(x_3)dl^\mu(x_3)]\,\ldots$$

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. I.e., this is a non-Abelian analogue of Aharonov-Bohm effect!

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

$$a_\mu \to U a_\mu U^{-1} - i\frac{\hbar}{q} U \partial_\mu U^{-1} \tag{5.12}$$

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. 5.5 (take $U = e^{i\chi}$). To see that this gauge transformation leaves the Wilson loop operators invariant (and hence is the right way to define a gauge transformation!) see Appendix section 5.6.1.

With a_μ a matrix valued quantity, the Chern-Simons action is now written as

$$S_{CS} = \frac{\hbar k}{4\pi}\int_{\mathcal{M}} d^3x\ \epsilon^{\alpha\beta\gamma}\ \mathrm{Tr}\left[a_\alpha \partial_\beta a_\gamma - \frac{2i}{3} a_\alpha a_\beta a_\gamma\right] \tag{5.13}$$

[18] These are named for Ken Wilson, who won a Nobel Prize for his work on the renormalization group and critical phenomena. There is a legend that Wilson had very very few publications when he came up for tenure as a professor at Cornell. Only due to the strong recommendation of his senior colleague Hans Bethe (already an Nobel Laureate at the time) did he manage to keep his job. Bethe knew what Wilson had been working on, and vouched that it would be extremely important. His ground-breaking work on renormalization group was published the next year. Everything worked out for him in the end, but the strategy of not publishing is *not* recommended for young academics trying to get tenure.

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

The Chern-Simons action is actually not quite gauge invariant. If in Eq. 5.12 we use unitary matrices $U(x)$ which are "close" to the identity (i.e., the whole set of matrices can be continuously deformed to the identity everywhere) 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

$$U(x) = \exp(i\epsilon H(x)) = 1 + i\epsilon H(x) \tag{5.14}$$

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. 5.12 and then into the Chern-Simons action, we find that (to lowest order in ϵ) the action is indeed gauge invariant (See appendix 5.6.2). 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

$$S_{CS} \to S_{CS} + 2\pi k n \hbar$$

for some integer n which is some sort of winding number[19] 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 $Z(\mathcal{M})$ turns out to be a manifold invariant (see chapter appendix, section 5.6.3), meaning tha smooth deformations of space and time do not change its value.

The insertion of the Wilson loop operator into the path integral gives a knot invariant of the link L that the Wilson loop follows[20]. Often we will think about our link as being embedded in a simple manifold like the three sphere, which we denote as S^3 (see section 3.1.1 for definition of S^3).

So for example, to find the link invariant corresponding to the two linked strings in Fig. 5.10, we have[21]

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

with W_L being the Wilson loop operators as in Eq. 5.10. Indeed, if we choose to work with the gauge group $SU(2)$ at level k we obtain the Kauffman invariant of the knot with $A = -(-i)^{(k+1)/(k+1)}$.

[19] In the case of the gauge group being $SU(2)$, as mentioned in section 3.2.2, 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. 5.14 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. A mathematician would say that $\Pi_3(S^3) = \mathbb{Z}$, meaning one can wrap S^3 around S^3 any integer number of times.

Fig. 5.10 A cartoon of a 3 manifold with a link embedded in it.

[20] 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. (See section 3.4) In field theory language this enters the calculation by how a point-splitting regularization is implemented.

[21] As mentioned by Eq. 5.10 we can choose to evaluate the trace of the Wilson loop operator in any representation of the gauge group. Choose different representations corresponds to world lines of different "particle types" in the theory!

5.6 Appendix: Odds and Ends about Chern Simons Theory

5.6.1 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.

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

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

$$W_C(x,x+dx) = 1 + i\frac{q}{\hbar}a_\mu dx^\mu \tag{5.15}$$

and its transformation should be

$$\begin{aligned} W_C(x,x+dx) &\to U(x)W_C(x,x+dx)U(x+dx)^{-1} \\ &= U(x)[1+i\frac{q}{\hbar}a_\mu dx^\mu]U(x+dx)^{-1} \\ &= U(x)[1+i\frac{q}{\hbar}a_\mu dx^\mu][U(x)^{-1}+dx^\mu\partial_\mu U(x)^{-1}] \\ &= 1+i\frac{q}{\hbar}[Ua_\mu U^{-1} - i\frac{\hbar}{q}U\partial_\mu U^{-1}]dx^\mu \end{aligned} \tag{5.16}$$

By comparing Eq. 5.15 and Eq. 5.16 we see that the gauge transform rule Eq. 5.12 correctly gives a gauge invariant Wilson loop operator.

5.6.2 Gauge Invariance of Chern-Simons action under small transforms

We want to use the small gauge transform Eq. 5.14 and expand everything to lowest order in ϵ. First, we plut this into Eq. 5.12 to obtain (setting $q = \hbar = 1$)

$$a_\mu \to a_\mu + \epsilon\,(i[H(x), a_\mu] - \partial_\mu H)$$

Plugging this into Eq. 5.13 and keeping terms linear in ϵ only one finds that all terms linear in ϵ vanish.[22]

[22] ***I need to track down why we have extra factor of q and $\hbar$ here, and check consistency of signs ***. Possibly a q hiding in Eq. 5.13)

5.6.3 Chern Simons Action is Metric Independent

You will often see books state that you don't see the metric $g_{\mu\nu}$ written anywhere in Eq. 5.13, therefore it must be metric independent. 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

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

But notice that the prefactor, including the ϵ, 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 and thus depends only on the topological properties of the manifold.

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!

5.6.4 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 sensitive 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_μ 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[23]. 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 theory, thus giving an end result of $|Z(\mathcal{M})|^2$ which obvioulsy won't care about the phase anyway!

[23] This is precisely what happens on the surface of materials known as "Topological Insulators" (or TIs) in three dimensions. The bulk of the system is a gapped insulator, but the surface of the system has a single Dirac fermion (or an odd number of Dirac fermions) and this is impossible to have in a purely two-dimensional system.

5.6.5 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

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

where here j indicates we have a loop around either cycle 1 (L_1) or cycle 2 (L_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 in section 5.3 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 Summary

- The Charge-Flux composite model describes abelian anyons — with the braiding phase coming from Aharonov-Bohm effect.
- We introduced idea of fusion, anitparticles and spin
- The vacuum for a system of anyons is nontrivial and can be a quantum memory.
- The Charge-Flux model can be realized in an abelian Chern-Simons theory.
- We introduced some ideas of general non-Abelian Chern-Simons theory, including manifold invariants and turning Wilson loop operators into knot invariants.

Further Reading

A good reference for the charge-flux composite model is John Preskill's lecture notes

- John Preskill, *Lecture Notes on Topological Quantum Computation*, http://www.theory.caltech.edu/p̃reskill/ph219/topological.pdf

A good reference for Abelian Chern-Simons theory is

- F. Wilczek, ed. *Fractional Statistics and Anyon Superconductivity*, World Scientific, (1990).

Some good references on Chern-Simons theory are

- *Current Algebras and Anomolies*, by S. Treiman, R. Jackiw, B. Zumino, and E. Witten (World Scientific) 1985. See particularly the chapters by R. Jackiw.
- E. Witten, *Quantum Field Theory and the Jones Polynomial* Comm. Math. Phys. Volume 121, Number 3 (1989), 351-399; available online here https://projecteuclid.org/euclid.cmp/1104178138. This is the paper that won a Fields' medal!
- Chetan Nayak, Steven H. Simon, Ady Stern, Michael Freedman, Sankar Das Sarma, *Non Abelian Anyons and Topological Quantum Computation*, Rev. Mod. Phys. 80, 1083 (2008). Also available online at https://arxiv.org/abs/0707.1889. This has a short discussion of Chern-Simons theory meant to be easily digested.
- Louis Kauffman, *Knots and Physics*, World Scientific, (2001), 3ed. The section on Chern-Simons theory is heuristic, but very useful.

Exercises

Exercise 5.1 Abelian Anyon Vacuum on a Two-Handle Torus
Using similar technique as in section 5.3, show that the ground state vacuum degeneracy on a two handle torus is m^2 for a system of abelian anyons with statistical angle is $\theta = \pi p/m$ for integers p and m relatively prime. Hint: Conside what the independent cycles are on a two-handled torus and determine the commutation relations are for operators T_i that take anyon-antianyon pairs around these cycles.

Short Digression on Quantum Gravity

6

6.0.1 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: 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 0.03 eV. Chemistry, visible light, and biology occur on the scale of 1 eV. 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!

[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.

6.0.2 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 am enthusiastic about.

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

6.1 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 15 orders of magnitude away from anything we have 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

$$Z = \sum_{\text{All universes}} e^{iS/\hbar} \tag{6.1}$$

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) Let us 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

[3]Written down first by Hilbert in 1915.

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

where the integration is over the entire space-time manifold $\mathcal{M}$, where here g is the space-time metric and R is the Ricci scalar. One might imagine that we could construct a theory of quantum gravity by plugging the Einstein-Hilbert action into the path integral form of Eq.6.1. 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 — these fluctuations are just gravity waves[4]. Even in this limit no one has fully made sense of this type of path integral without many additional assumptions.

[4]Observation of gravity waves by the LIGO experiment won the 2017 Nobel prize.

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

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, section 6.2.) 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. 6.1 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 $\mathcal{M}$.

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 shown in Fig. 6.1.

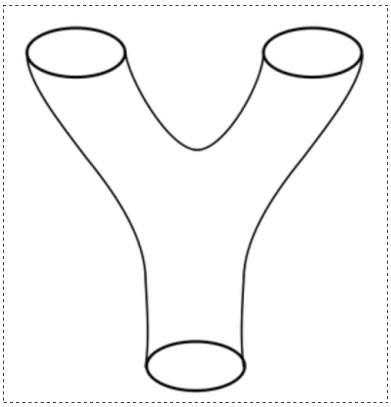


Fig. 6.1 A manifold where the topology of a time-like slice changes as time progresses. Time runs vertically in this picture

OK, here is the surprise: the function $Z(\mathcal{M})$ is precisely the Chern-Simons partition function discussed above in section 5.5 for an appropriately chosen gauge group!

6.2 Appendix: No Gravity Waves in 2+1 D

Why are there no gravity waves in 2+1 dimension? The short argument for this is as follows (taken from the article by S. 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.

Let us 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 $\epsilon_{\mu\nu}$ is orthogonal to the lightlike propagation wavevector, $k^\mu k_\mu = 0$ meaning

$$\epsilon_{\mu\nu} k^\nu = 0. \tag{6.2}$$

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. 6.2 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. 6.2 is 3 constraints, and A has three parameters, so we should always be able to solve the equation for A given ϵ.

Further Reading

- For a huge amount of information on 2+1 dimensional quantum gravity, see S. Carlip, http://www.livingreviews.org/lrr-2005-1.
- The relationship of 2+1 D gravity to Chern-Simons theory was first developed in A. Achucarro and P. Townsend "A Chern-Simons Action for Three-Dimensional anti-De Sitter Supergravity Theories". Phys. Lett. B180: 89 (1986).
- The relationship was further developed in Edward Witten, "(2+1)-Dimensional Gravity as an Exactly Soluble System," Nuclear Physics B. 311 (1): 4678 (1988).
- Years later, the question was revisited by the same author in Edward Witten "Three-Dimensional Gravity Revisited", arXiv:0706.3359.

Topological Quantum Field Theory

7

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, but is often not very useful.

Another way to define a (2+1 dimensional) TQFT is that it is a set of rules that takes an input of a labeled link embedded in a three-manifold and gives an output of a complex number in a way that is invariant under smooth deformations. This definition is quite analogous to our definition of a knot invariant, with two key differences. First, we allow for the lines to be labeled with a "particle type" (and our rules for evaluating the end result will depend on the particular particle type labels). Secondly, the link can be embedded in some arbitrarily complicated three-manifold This type of mapping (see Fig. 7.1) is precisely the sort of thing that one gets as an output of Chern-Simons theory which we called $Z(M, \text{links})$ as we discussed in section 5.5. The advantage of thinking in this language is that strictly speaking, the functional integrals of Chern-Simons theory are often not well defined mathematically.

A closely related but more formal definition of TQFTs is given by a set of Axioms by Sir Michael Atiyah[1] which are in some sense much more informative.

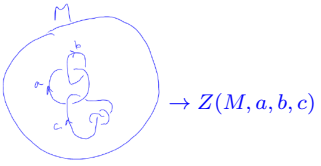


Fig. 7.1 A (2+1) dimensional TQFT takes an input of a labeled link in a manifold and produces an output of a complex number in a manner which is topologically invariant.

[1] Sir Michael Atiyah, a Fields medalist, was one of the foremost mathematicians of the 20th century. He specialized in geometry and topology — particularly at the interface between mathematics and physics.

7.1 Paraphrasing of Atiyah's Axioms

Here I'm going to give a rough interpretation of Aityah's axioms of TQFT, suitable for physicists. To begin with, we will consider space-time manifolds with no particles in them. As we have found above, TQFTs are nontrivial even in the absence of any particles. 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[2] which we call $\mathcal{M}$, and d dimensional oriented slice Σ — we can often think of this

[2] While it is possible to define certain TQFTs on non-orientable manifolds it is much easier to assume that all manifolds will be orientable — excluding things like Möbius strips and Klein bottles.

slice as being the d-dimensional 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[3] of Σ.

[3]The phrases "depends only on the topology..." is something that physicists use in this sense, but mathematicians would not (topology can describe things like whether sets are closed or open, whether points are infinitely dense, etc.). Perhaps it would be better to just say that $V(\Sigma)$ does not change under continuous deformation of Σ.

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

As an example of what we mean, we have seen that if Σ is a torus, there is a nontrivial Hilbert space coming from the ground state degeneracy. This degenerate space is the space $V(\Sigma)$. The space $V(\Sigma)$ will depending on the particular anyon theory we are considering (for example in the case of Abelian anyons we found the degeneracy for a system with statistical angle $\theta = \pi p/m$ is m).

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 also 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. Let us state this mathematically.

Axiom 2 Implies:

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

[4]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.

The reason this must be true is because $\emptyset \cup \Sigma = \Sigma$ and $\mathbb{C} \otimes V(\Sigma) = V(\Sigma)$ so the result follows[4].

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 particular 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 $\mathcal{M}$ (the interior, not the boundary) is the space-time history of the system, and $Z(\mathcal{M})$ is the particular wavefunction that is picked out by this given space-time history (See Fig. 7.2).

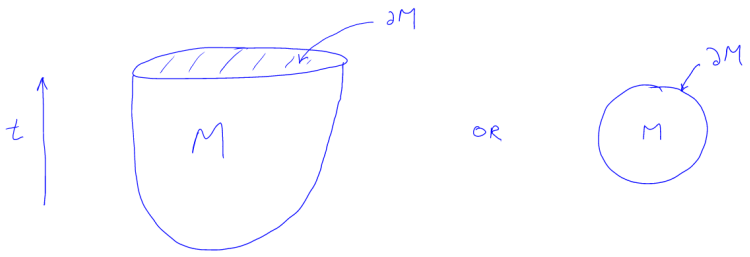


Fig. 7.2 Two depictions of a space-time manifold $\mathcal{M}$ with boundary $\partial\mathcal{M}$. The left depiction is problematic because the only boundary of the manifold is supposed to be the top surface ∂M. The right depiction is more accurate in this sense, although it depicts a 2D $\mathcal{M}$ 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.

Axiom 3 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. 7.3.

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. 7.4. Here we have $\Sigma = \partial\mathcal{M} = (\partial\mathcal{M}')^*$ so we can glue together the two manifolds along their common boundary to give[5]

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

[5]The notation $\mathcal{M} \cup_\Sigma \mathcal{M}'$ means the union of $\mathcal{M}$ and $\mathcal{M}'$ glued together along the common boundary $\cup_\Sigma$

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

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

We say that $\mathcal{M}$ is a cobordism between Σ_1 and Σ_2. See Fig. 7.5 for an example.

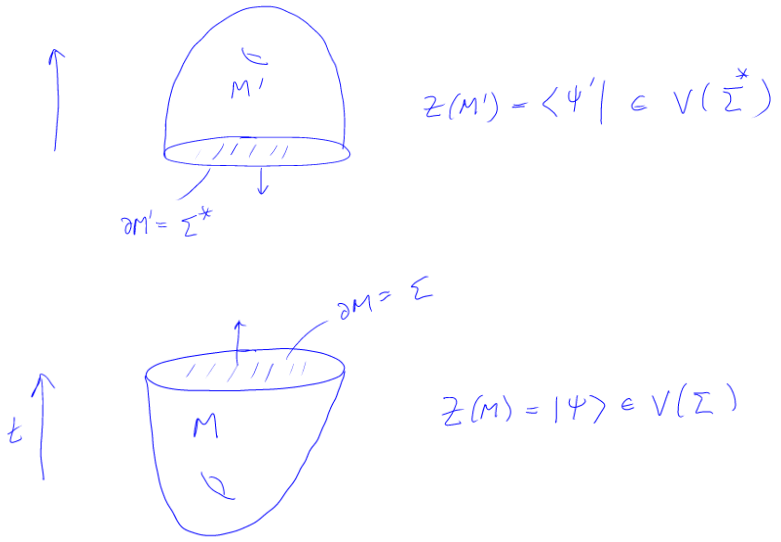


Fig. 7.3 In this picture $\mathcal{M}$ and $\mathcal{M}'$ are meant to fit together since they have a common boundary but with opposite orientation $\Sigma = \partial\mathcal{M} = \partial\mathcal{M}'^*$. Here $\langle\psi'| = Z(\mathcal{M}') \in V(\Sigma^*)$ lives in the dual space of $|\psi\rangle = Z(\mathcal{M}) \in V(\Sigma)$

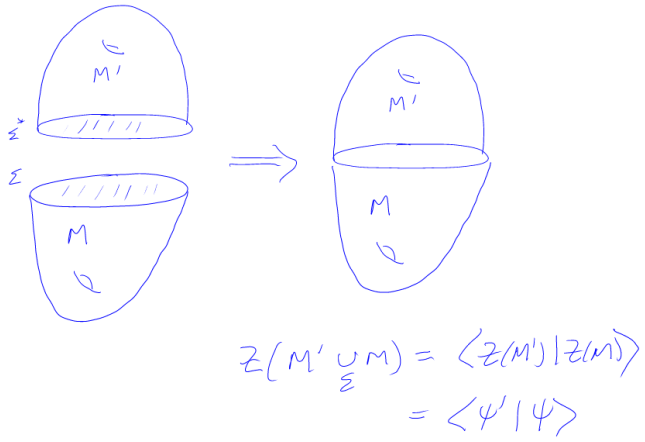


Fig. 7.4 Gluing two manifolds together by taking the inner product of the wave-functions on their common, but oppositely oriented, boundaries.

We thus have $Z(\mathcal{M}) \in V(\Sigma_1^*) \otimes V(\Sigma_2)$, so that we can write

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

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 evolution[6] between the vector spaces $V(\Sigma_1)$ and $V(\Sigma_2)$.

[6]This evolution may or may not be unitary — indeed, the dimensions of $V(\Sigma_1)$ and $V(\Sigma_2)$ may not even match if $\Sigma_1 \neq \Sigma_2$. For example, capping off a cyclinder with a half-sphere, must act as a projection, which is not unitary.

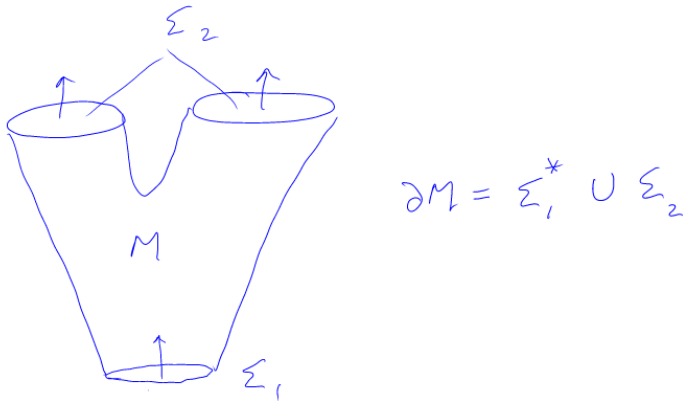


Fig. 7.5 $\mathcal{M}$ is the cobordism between Σ_1^* and Σ_2. I.e., $\partial\mathcal{M} = \Sigma_1^* \cup \Sigma_2$. Note that we have reversed orientation of Σ_1 here.

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) then the boundaries are Σ and Σ^* (See Fig.7.6), 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. 7.6.

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

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 Fig. 7.6. We then have

$$Z(\Sigma \times S^1) = \text{Tr}\left[\sum_\alpha |\psi_{\Sigma,\alpha}\rangle \otimes \langle\psi_{\Sigma,\alpha}|\right] = \text{Dim}[V(\Sigma)]. \tag{7.1}$$

Thus we 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 in section 5.3, for the torus T^2 we have

$$\text{Dim}\, V(T^2) = \#\text{ of particle species} \tag{7.2}$$

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

$$\text{Dim}\, V(S^2) = 1 \tag{7.3}$$

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

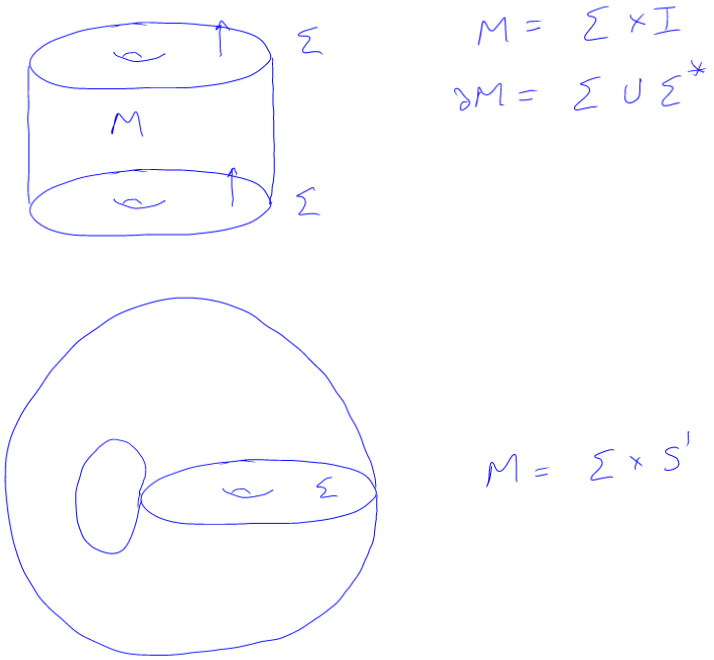


Fig. 7.6 Top: A cobordism that can be topologically contracted to nonthing acts as the identity on the Hilbert space $V(\Sigma)$. Bottom: Gluing the top of $\Sigma \times I$ to the bottom we obtain $\mathcal{M} = \Sigma \times S^1$. The important fact is that $Z(\Sigma \times S^1)$ is just the ground state degeneracy of Σ as a 2-manifold.

7.2 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.

Let us imagine that 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. 7.7. 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 (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

Fig. 7.7 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.

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. 7.8. 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

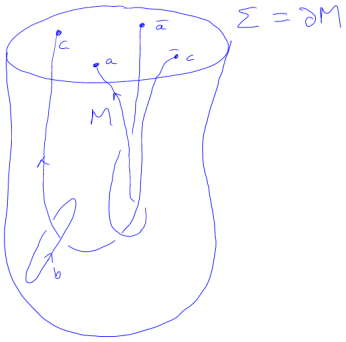


Fig. 7.8 A 3-manifold $\mathcal{M}$ with particles in it, which are marked and labeled lines (the lines should be directed unless the particle is its own antiparticle). These world lines may end on the boundary $\Sigma = \partial\mathcal{M}$. The wavefunction on the boundary $\partial\mathcal{M}$ is 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. 7.9. The final wavefunction 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

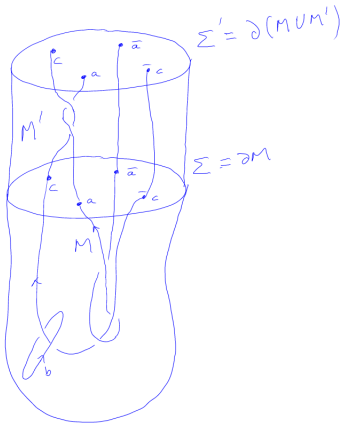


Fig. 7.9 $\mathcal{M}'$ evolves the positions of the particles in time. Note that by $\mathcal{M}'$ we mean not just the manifold, but the manifold along with the world-lines in it.

statistics. We can write explicitly the unitary transformation

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

Note that if the particles stay fixed in their positions (or move in topologically trivial ways) then $\mathcal{M}'$ can be contracted to nothing 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

$$Z(\Sigma \times S^1) = \mathrm{Dim}[V(\Sigma)] \tag{7.4}$$

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.

7.2.1 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. 7.10. 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 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

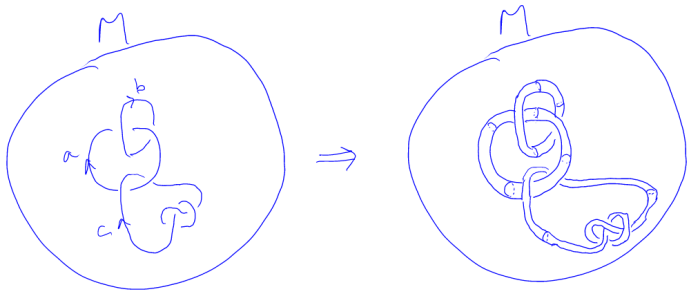


Fig. 7.10 Removing the world-lines on the left along with a thickened tube. Imagine a worm burrowing along the path of the world lines and leaving a hollow hole.

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 Fig. 7.11, where a particle of type a goes around the handle. Obviously, gluing such a solid torus back into the empty solid-torus-shaped tube recovers the original picture of labeled world lines following these paths. We can think of this solid torus

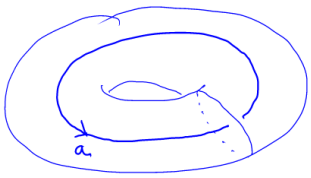


Fig. 7.11 The possible wavefunctions $|\psi_a\rangle$ that we can have on the surface of the torus can be realized by having a world-line of a particle of type a going around the handle of the torus. We can call these $Z(\text{solid torus with } a \text{ running around handle}) = |\psi_a\rangle$

manifold as being a space-time history where $t = -\infty$ is the central core of the solid torus (the circle that traces the central line of the jelly filling of the donut) and the torus surface is the present time. Somewhere between $t = -\infty$ and the time on the surface of the torus, a particle of type a has been dragged around the handle.

The manifold with the tori excised from it (the right of Fig. 7.10) contains all of the information necessary to give a partition function for the left of Fig. 7.10 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 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 $|\psi_{L1,a}\rangle$, $|\psi_{L2,b}\rangle$, $|\psi_{L3,c}\rangle$.

7.3 Building 3-Manifolds

7.3.1 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[7]

$$S^3 = (S^1 \times D^2) \cup_{T^2} (D^2 \times S^1)$$

The notation here is that the two pieces $S^1 \times D^2$ and $D^2 \times S^1$ are joined together on their common boundary which is T^2 (the torus surface).

There is a very elegant proof of this decomposition. Consider the 4-ball D^4. Topologically we have[8]

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

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

$$S^3 = \partial B^4 = \partial(D^2 \times D^2) = (S^1 \times D^2) \cup_{T^2} (D^2 \times S^1)$$

where we have used the fact that the boundary of a disk is a circle, $\partial D^2 = S^1$. Note that the two solid tori 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 two tori 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. 7.12.

[7] If you are rusty on these elementary topology manipulations, see the review in section 3.1

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

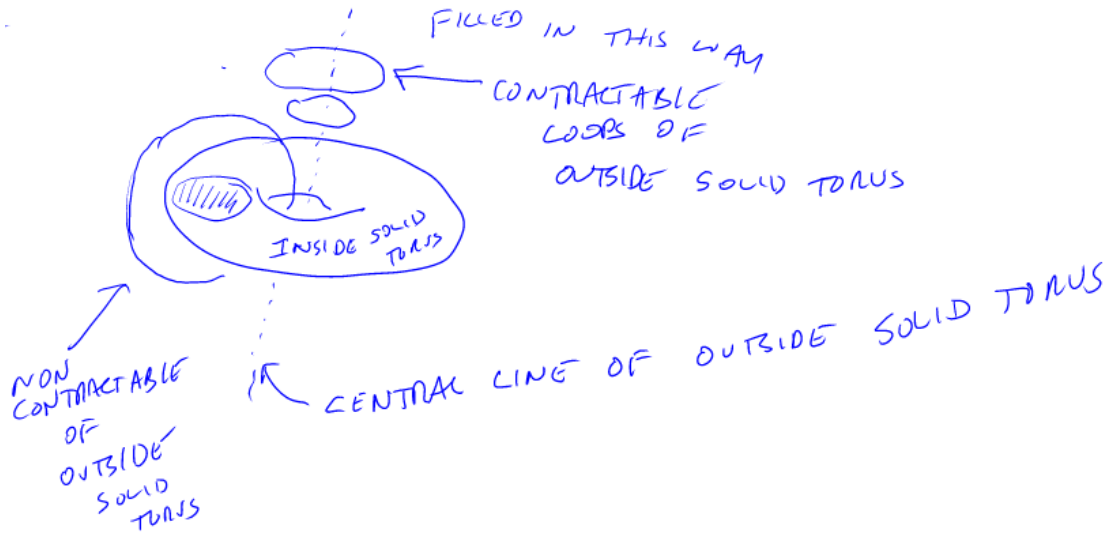


Fig. 7.12 Assembling two solid toruses to make S^3. One torus (labeled "inside solid torus") is obvious. Here the contractable disk D^2 is shaded and the noncontractable S^1 is the handle. The remainder of space, including the point at infinity is the other "outside" solid torus. For the "outside" solid 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.

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 inside tori.

$$Z(S^3) = \langle Z(D^2 \times S^1) | Z(S^1 \times D^2) \rangle = \langle \psi_{inside} | \psi_{outside} \rangle$$

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

We can further consider including world lines around the noncontractable loops of the solid torus, as in Fig. 7.11. There is a different state on the surface of the torus for each particle type we 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. 7.13. Gluing the two tori together we get

$$Z(S^3; a \text{ loop linking } b \text{ loop}) = \langle Z(D^2 \times S^1; a) | Z(S^1 \times D^2; b) \rangle \equiv S_{ab} \tag{7.5}$$

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.[9]

Note that the S-matrix is unitary, since it is simply giving a basis transformation between the two sets of wavefunction which both span

[9] Some comments on the S-matrix: (1) since a linking b is topologically the same as b linking a we should have $S_{ab} = S_{ba}$. (2) Reversing the direction of the world line takes a particle to its anti-particle. This is topologically the same as taking the mirror image of the linking diagram in Fig. 7.13, thus we have $S_{\bar{a}b} = [S_{ab}]^*$ where $\bar{b}$ is the antiparticle of b.

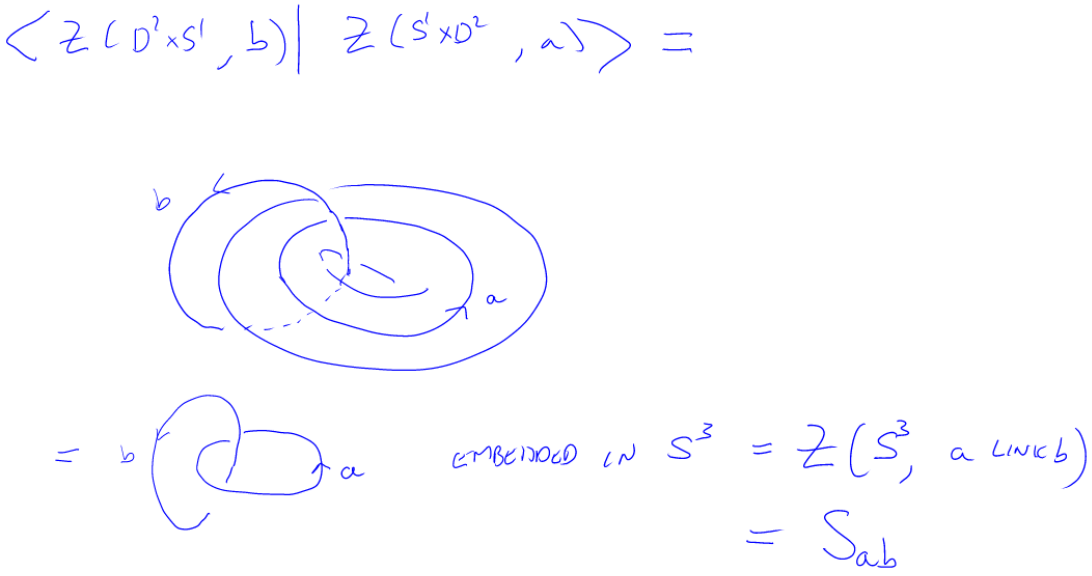


Fig. 7.13 Here we assemble a partition function for S^3 with world lines of a linking b embedded in the S^3. To do this we glue together two solid tori each with a world line running around the handle. The end result is known as the modular S-matrix, and it gives a basis transform converting between the two bases which both span the hilbert space of the torus surface where the two solid tori are glued together.

the vector space $V(T^2)$ of the torus surface T^2 where the two solid toruses are glued together.

Note that the element S_{00}, corresponding the element of the S-matrix where the vacuum particle (no particle at all!) is put around both handles. (Here we are using 0 to mean the vacuum.) This tells us that

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

and in fact, should be strictly less than one unless there are no nontrivial particle types.

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.

7.3.2 $S^2 \times S^1$

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 7.14. Here we want to show that[10]

[10] One should be warned that $S^2 \times S^1$ cannot be embedded in usual three dimensional space, so visualizing it is very hard!

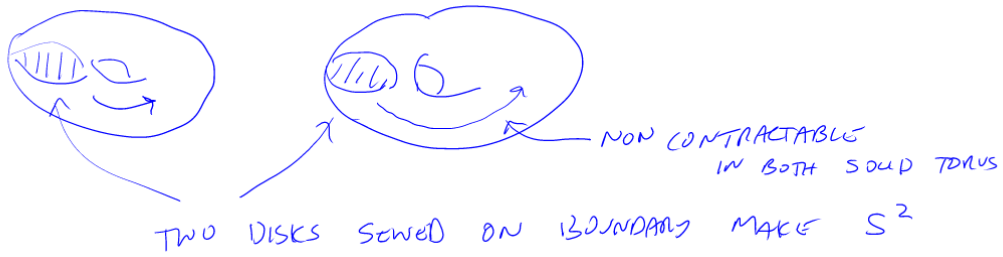


Fig. 7.14 Assembling two solid tori to make $S^2 \times S^1$. Here the two contractable disks D^2 are sewed together along their boundaries to make S^2.

$$S^2 \times S^1 = (D^2 \times S^1) \cup_{T^2} (D^2 \times S^1)$$

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 solid tori if we want. If we do so 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 two of our prior statements. On the one hand Eq. 7.4 says that Z for any manifold $\times S^1$ should be the dimension of the Hilbert space for that manifold; and on the other hand Eq. 7.3 states that the dimension of the Hilbert space on a sphere is 1.

7.4 Surgery and More Complicated Manifolds

The understanding of 3-manifolds is a very difficult problem[11]. 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 a manifold by gluing pieces together — the idea of surgery is that we

[11] Many important results on three manifolds have been discovered recently. Perelman's proof of Poincare Conjecture, along with the methods he used are apparently extremely powerful. But this is *way* outside the scope of our book!

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. We split a closed manifold $\mathcal{M}$ into two pieces $\mathcal{M}_1$ 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 $\mathcal{M}'$.

7.4.1 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. 7.15. 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 $S^1 \cup S^1$.

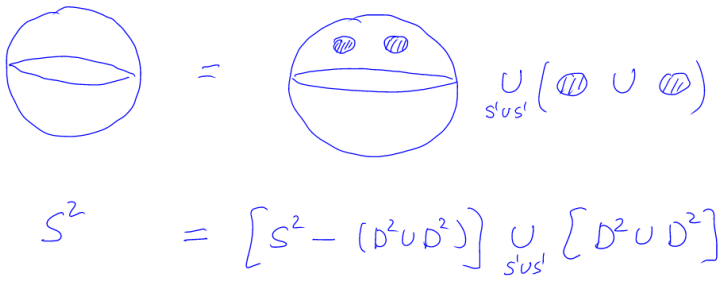


Fig. 7.15 Writing a sphere $\mathcal{M} = S^2$ as the union of two manifolds glued along their boundaries. $\mathcal{M}_2$ is the union of two disks $D^2 \cup D^2$. $\mathcal{M}_1 = S^2 - (D^2 \cup D^2)$ is the remainder. The two manifolds are glued along their common boundary $S^1 \cup S^1$.

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 removed $\mathcal{M}_2$, as shown in Fig. 7.16. The resulting manifold $\mathcal{M}'$ is the torus T^2

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

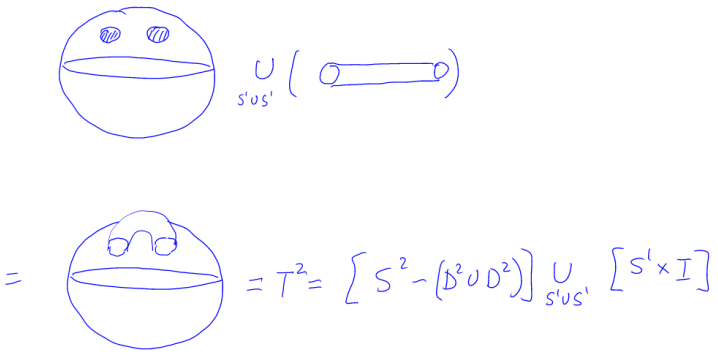


Fig. 7.16 Gluing the cylinder surface $\mathcal{M}_2' = S^1 \times I$ to the manifold $\mathcal{M}_1 = S^2 - (D^2 \cup D^2)$ along their common boundary $S^1 \cup S^1$ gives the torus T^2.

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 $\mathcal{N} = B^3$ the 3-ball (sometimes denoted D^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. 7.17

$$\begin{array}{ccc} \mathcal{N} = B^3 & & \partial\mathcal{N} = \mathcal{M} = S^2 \\ \downarrow \text{ Add Handle} & & \downarrow \text{ Surgery} \\ \text{Solid Torus} & & \partial(\text{Solid Torus}) = T^2 \end{array}$$

Fig. 7.17 Handle attaching on the manifold $\mathcal{N}$ is the same as surgery on a manifold $\mathcal{M} = \partial\mathcal{N}$.

7.4.2 Surgery on 3-manifolds

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 the pieces of your exploded head.) The reason this is difficult to visualize is 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 solid 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 solid 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. 7.10.

Lickorish-Wallace Theorem

This is quite an important theorem of topology.[12]

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

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).

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 solid torus. Excise each of these solid tori, and replace them by tori with longitude and meridian switched. Any possible 3-manifold can be obtained in this way by surgering an appropriately chosen link. We summarize with

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

We can thus represent any three dimensional manifold in terms of a link. If we think of a topological quantum field theory as being a way to assign a complex number to a three dimensional manifold, i.e., $Z(\mathcal{M})$ we realize that what we are now looking for is essentially a knot invariant — a way to assign a number to a knot. We explore this connection further when we discuss Witten-Reshitikhin-Turaev invariant below in this section.

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. 7.18.

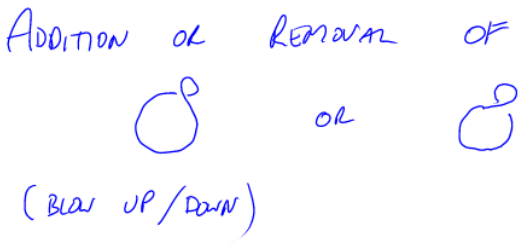


Fig. 7.18 Blow up/ Blow down.

(2) **Handle-Slide** A string can be broken open and pulled along the full path of another string, and then reconnected. See Fig. 7.19.

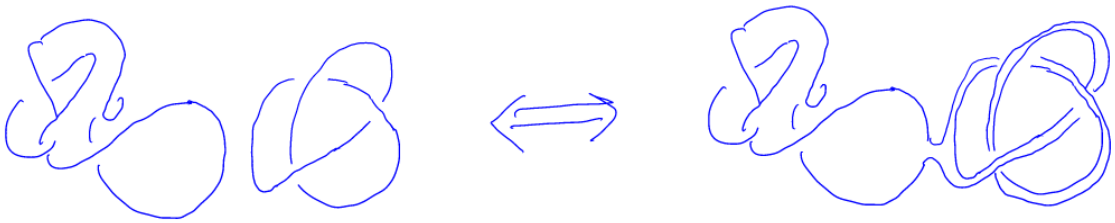


Fig. 7.19 A handle-slide move

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.

Witten-Reshitikhin-Turaev Invariant

Here we are interested in constructing what is known as a manifold invariant. Similar to a knot invariant, this is a mapping from a manifold to some output that depends only on the topological properties of the input manifold.

Our strategy of building a manifold invarient is to describe the manifold by using surgery on a link. Given knowledge of the rules of Kirby calculus, to construct a manifold invariant for three manifolds, we 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 the manifold invariants of Chern-Simons theory mathematically rigorous.

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). Thus each anyon theory gives us a way (many ways, actually) to construct knot invariants. 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 (We will show details of this later in ***). The second Kirby move seems harder to achieve, but can be achieved if one uses the so-called Kirby color combination (sometimes known as an 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, and the subscript 0 refers to the vacuum or identity particle. Diagrammatically we have Fig. 7.20. It turns out (See

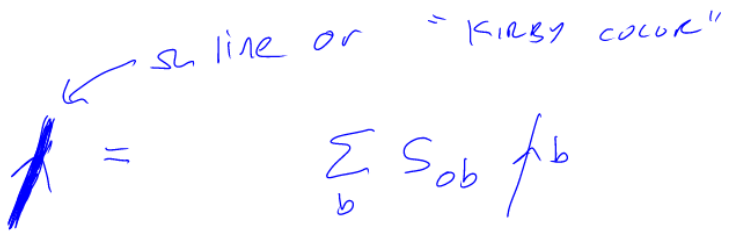


Fig. 7.20 A String of Kirby Color is a weighted superposition of all anyon string types

exercise 12.2) 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. (See chapter *** for more details).

Further Reading

For discussion on the Atiya Axioms

- M. F. Atiyah, *Proceedings of 5th Gokova Geometry and Topology Conference*, Tr. J. Mathematics, 21, 1, (1997). http://www.maths.ed.ac.uk/ aar/papers/atiyahinttqft.pdf

- M. F. Atiyah, *Topological quantum field theory.* Publications Mathmatiques de l'IHS, 68 (1988), p. 175-186 http://www.numdam.org/item?id=PMIHES_1988__68__175_0

For more detailed discussion of Surgery and Kirby Calculus, see

- Robert E. Gompf and Andr/'as I. Stipsicz, *4-Manifolds and Kirby Calculus*, American Mathematical Society, Graduate Studies in Mathematics Volume: 20 (1999).
- Robion Kirby, *The Topology of 4-Manifolds*, Springer (1989).

Exercises

Exercise 7.1 Surgery

(a) Beginning with the three-sphere S^3, consider the so-called "unknot" (a simple unknotted circle S^1 with no twists) embedded in this S^3. Thicken the circle into a solid torus ($S^1 \times D^2$) which has boundary $S^1 \times S^1$. Now perform surgery on this torus by excising the solid torus from the manifold S^3 and replacing it with another solid torus that has the longitude and meridian switched. I.e., replace $S^1 \times D^2$ with $D^2 \times S^1$. Note that both of the two solid tori have the same boundary $S^1 \times S^1$ so that the new torus can be smoothly sewed back in where the old one was removed. What is the new manifold you obtain? (This should be easy because it is in the book!)

(b) [Not hard if you think about it right!] Consider two linked rings, known as the Hopf link (See Fig. 7.21). Consider starting with S^3 and embedding the

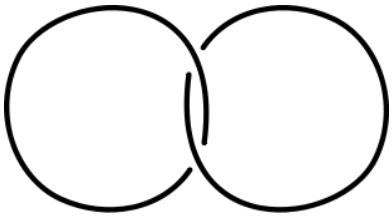


Fig. 7.21 Hopf Link

Hopf link within the S^3 with "blackboard framing" (i.e., don't introduce any additional twists when you embed it). Thicken both strands into solid tori and perform surgery on each of the two links exactly as we did above. Argue that the resulting manifold is S^3.

(c) [Hard] Consider the link shown in Fig. 7.22 known as the Borromean rings. This is an interesting link because no two strands are actually linked with each other, but the three links are still tied together. If you remove any one strand the remaining two come apart.

Consider starting with S^3 and embedding the Borromean rings within the S^3 with "blackboard framing". Thicken all three strands into solid tori and perform surgery on each of the three links exactly as we did above. What manifold do you obtain? Hint 1: Think about the group of topologically different loops through the manifold starting and ending at the same point,

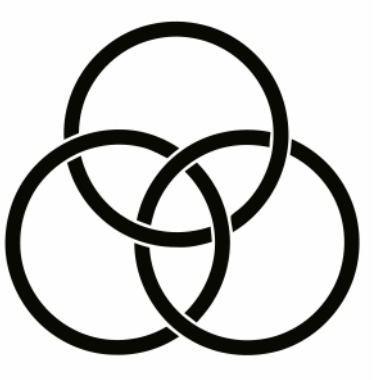


Fig. 7.22 Borromean Rings

the so-called "fundamental group" or first homotopy group. (See section 3.3). Hint 2: If we say a path around the meridian of one of the three Borromean rings (i.e., threading though the loop) is called a and the path around the meridian of the second ring is called b, then notice that the third ring is topologiclly equivalent to $aba^{-1}b^{-1}$.

Fusion and Structure of Hilbert Space

8

As discussed above, each two-dimensional surface (a slice of a three-dimensional space-time manifold) has an associated 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 this Hilbert space and how it depends on the particles.

8.1 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 which we call fusion, or do the reverse, which we call 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. 8.1. These diagram

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. 8.2.

For example, in our abelian anyon model of charges and fluxes (see section 5.2), if the statistical angle is $\theta = \pi p/m$ (p and m relatively prime and not both odd) then we have m species $a = (aq, a\Phi)$ for $a = 0 \ldots m-1$, where $q\Phi = \pi p/m$. The fusion rules are simply addition modulo m. That is $a \times b = (a+b) \bmod m$.

Identity

Exactly one of the particles should be called the identity or vacuum. We write this[1] as 1 or 0 or I or e. The identity fuses trivially

$$a \times I = a$$

Fig. 8.1 Fusion and splitting diagrams can be thought of as part of a space-time history of the particles. If we are describing two separated particles a and b whose overall quantum number is c, we would describe the ket for this state using the right hand picture — which we can think of as a space-time description of how the current situation came about (with time going up). Conversely, the left picture can be thought of as the corresponding bra. More details of the relationship of the diagrammatic algebra to kets are given in chapter 9.

Fig. 8.2 Another notation to describe th fusion of two particle types to make a third $a \times b = c$. The two particles need not be close to each other.

[1] It is admittedly annoying that we have so many different ways to express the identity, but in different contexts different notations seem natural. For example, if our group of particles is fusing by addition (as we discussed in the charge-flux model) the identity should be 0. But if our group fuses by multiplication, identity is more naturally 1.

Fig. 8.3 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.

Fig. 8.4 A particle going forward should be equivalent to an antiparticle going backwards.

Fig. 8.5 Fusion of an anyon with its anti-anyon to form the identity can be thought of as a particle turning around in space-time.

Fig. 8.6 Multiple possible fusion channels

for any particle type a. In the flux-charge model (section 5.2) we should think of the identity as being no charge and no flux. Fusion with the identity is depicted schematically in Fig. 8.3

Antiparticles

Each particle a should have a unique antiparticle which we denote as $\bar{a}$. The antiparticle is defined by $a \times \bar{a} = I$. (There should only be one particle which fuses with any a to give the identity!). A particle going forward should be equivalent to an antiparticle going backwards as shown in Fig. 8.4. Fusion to the identity can be thought of as particle turning around in space-time as shown in Fig. 8.5.

A particle may be its own antiparticle, 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 $\theta = \pi p/m$.

8.2 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. 8.6. 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). 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 $V(\Sigma)$).

What is this Hilbert space associated with multiple fusion channels? 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 of angular momentum is the identity here, it has no spin at all. The analogy with 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 predominant theme of anyon physics (if not of physics altogether).

The quantum number (or "charge") of a particle is locally conserved in space. Consider, for example, Fig. 8.7. On the left, a particle a propagates 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 Fig. 8.1). We sometimes call this the **no transmutation** principle. It allows us to conclude that the complicated picture on the left of Fig. 8.7 must be equal to some constant times the simple propagation of an a particle as shown on the right.

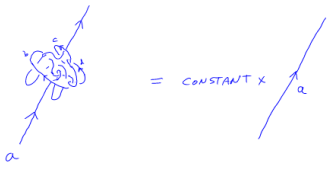


Fig. 8.7 If a particle a goes into a spacetime region, then a net particle charge a must come out. This is also sometimes called the "no-transmutation" principle. From far away, one can ignore any local processes (up to an overall constant).

If two particles (maybe far away from each other) fuse together to some particle type (in a case where multiple fusion channels are available) it is not possible to determine this fusion channel by measuring only one of the initial particles. In order to determine the fusion channel of the two particles, you have to do an experiment that involves both of the initial particles. For example, one can 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. 8.8), no amount of braiding b around c will change this overall fusion channel. The fusion channel is *local* to the pair. 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.

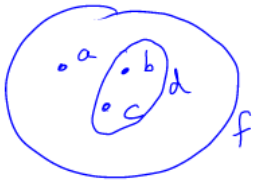


Fig. 8.8 In this picture b and c fuse to d. Then this d fuses with a to give an overall fusion channel of f. No amount of braiding b around c will change the fact that the two of them fuse to d. However, if we braid a with b and c, this can change the fusion of b with c subject to the constraint that the fusion of all three particles will give f.

Antiparticles in the Case of 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.

8.2.1 Example: Fibonacci Anyons

Perhaps the simplest nonabelian example is the anyon system known as Fibonacci[2] 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[3].

[3]Fibonacci anyons can be described exactly by the G_2 level 1 Chern-Simons theory. This involves a messy Lie algebra called G_2. The $SU(2)_3$ Chern-Simons theory

[2]Fibonacci's was born in Pisa around 1175 AD. Perhaps his most important contribution to mathematics is that brought Arabic numerals (or Hindu-Arabic numerals) to the western world. The Fibonacci sequence $1, 1, 2, 3, 5, 8, 13, \ldots$ is named after him, although it was known in India hundreds of years earlier!

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

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

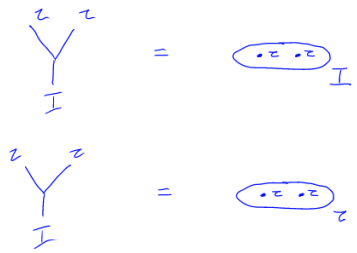


Fig. 8.9 Two different notations for the two different fusion channels of two Fibonacci anyons

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 this rule 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. 8.9.

With three fibonacci anyons the Hilbert space is 3 dimensional, as shown in Fig. 8.10. 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.

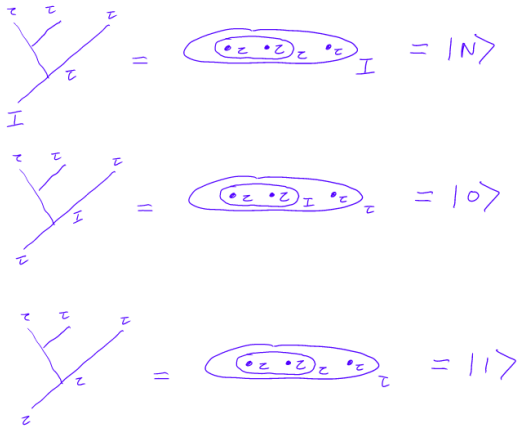


Fig. 8.10 Notations for the three different fusion channels of three Fibonacci anyons. The notation $|N\rangle, |0\rangle$ and $|1\rangle$ are common notations for those interested in topological quantum computing with Fibonacci anyons! Note that fusion trees have branches pointing upwards here which is the conventional notation for a ket (bras have branches pointing down. See chapter 9)

There are two states in the Hilbert space of three anyons (labeled $|0\rangle$ and $|1\rangle$ in Fig. 8.10) which both have an overall fusion channel of τ. As

contains some additional particles besides the Fibonacci particles, but ignoring these, it is the same as Fibonacci.

mentioned above in section 8.6, 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 in section ***, 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

$$\text{Dim of } N \text{ Anyons} = \text{Fib}_N \sim \left(\frac{1+\sqrt{5}}{2}\right)^N \tag{8.1}$$

we say that the *quantum dimension* of this particle is $d = (1+\sqrt{5})/2$, the golden mean (See Eq. 4.5).

8.2.2 Example: Ising Anyons

The Ising anyon system is extremely closely related to $SU(2)_2$ Chern-Simons theory[4], 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.

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

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 σ's 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 8.11.

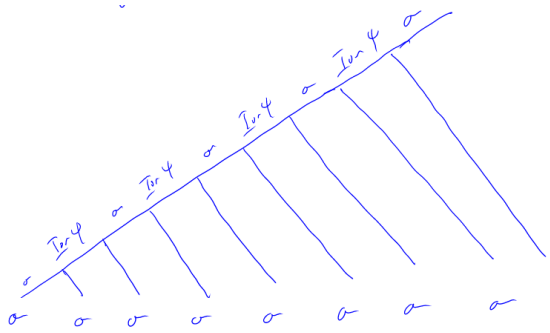


Fig. 8.11 The Ising Fusion Tree.

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 $(\sqrt{2})^N$ the quantum dimension of the σ particle is $d = \sqrt{2}$ (See Eq. 4.5).

8.3 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, with each particle having a unique inverse or antiparticle.

The general fusion rules can be written as

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

where the N's are known as 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[5] 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

[5]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.

most simple 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 $a \times b = b \times a$.

It is sometimes convenient to define

$$N_{ab\bar{c}} = N_{ab}^c \tag{8.2}$$

which is the number of different ways that a, b, and $\bar{c}$ can fuse to the identity. This equivalence is shown graphically in Fig. 8.12. The advantage of this representation is that N_{abc} is fully symmetric in all of its indices.

Fig. 8.12 Equivalence of N_{ab}^c with $N_{ab\bar{c}}$.

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 $\bar{a}$ meaning a particle is its own antiparticle.

If we are to fuse, say, five particles of type a together, we can do so via a tree as shown in Fig. 8.13. To find the dimension of the Hilbert space, we write

$$\begin{aligned} \text{Dimension of fusing five } a \text{ particles} &= \sum_{bcde} N_{aa}^b N_{ba}^c N_{ca}^d N_{da}^e \\ &= \sum_{bcde} N_{aa}^b N_{ab}^c N_{ac}^d N_{ad}^e \end{aligned}$$

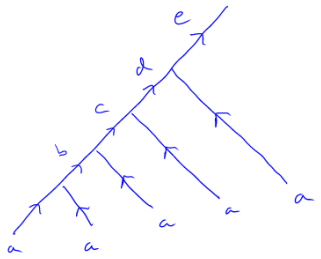


Fig. 8.13 Fusing five a particles together

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

We recall (See Eq. 4.5) that the quantum dimension d_a of the particle a is defined via the fact that the Hilbert space dimension should scale as d_a^N where N is the number of a particles fused together. To find this quantum dimension 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

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

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 $d_\tau = (1+\sqrt{5})/2$, in agreement with Eq. 8.1.

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 $d_\sigma = \sqrt{2}$ as described in section 8.2.2.

8.3.1 Associativity

It should be noted that the fusion multiplicity matrices N are pretty special matrices 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. 8.14 or we can try fusing b and c first as on the right. Correspondingly to these two possibilities we have the equality

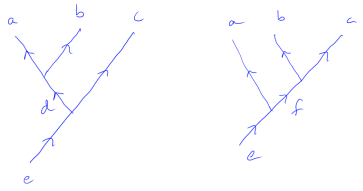


Fig. 8.14 Fusing $(a \times b) \times c$ should be equivalent to $a \times (b \times c)$

$$\sum_d N_{ab}^d N_{cd}^e = \sum_f N_{cb}^f N_{af}^e \tag{8.3}$$

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

$$[N_a, N_c] = 0 \tag{8.4}$$

Therefore all of the N matrices commute with each other. In addition the N's are *normal* matrices, meaning that they commute with their own transpose[6]. A set of normal matrices that all commute are simultaneously diagonalizable, thus

$$[U^\dagger N_a U]_{xy} = \delta_{xy} \lambda_x^{(a)} \tag{8.5}$$

and all N_a's get diagonalized with the same matrix U. Surprisingly (as we will see below in section ***) the matrix U is precisely the modular S-matrix we discussed above in Eq. 7.5 !

[6] To see this note that by time reversal we should have $[N_a]_b^c = [N_{\bar{a}}]_{\bar{b}}^{\bar{c}}$. But using Eq. 8.2, we have $[N_a]_b^c = [N_{\bar{a}}]_c^b$ or $N_a = N_{\bar{a}}^T$. But we know that all the N's commute, so N commutes with its own transpose.

8.4 Fusion and Hilbert Space

The structure of fusion rules can be used to calculate the ground state degeneracy of wavefunctions on 2-dimensional manifolds[7].

Let us start by considering the sphere S^2, and 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. All other configurations (change of topology, adding particles etc) will 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-antiparticles 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 particle with an $\bar{a}$ particle, 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 sphere, without labeling the particle type looks like punctured sphere in Fig.7.15.). In the spirit of what we did in section 7.2.1 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 can only get 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. 8.15. 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

[7] We are again assuming manifolds are always orientable – so this excludes objects like the Klein bottle or the Möbius strip.

described in Fig. 7.16. The result of this surgery is to give the torus surface T^2 and we conclude that we should have

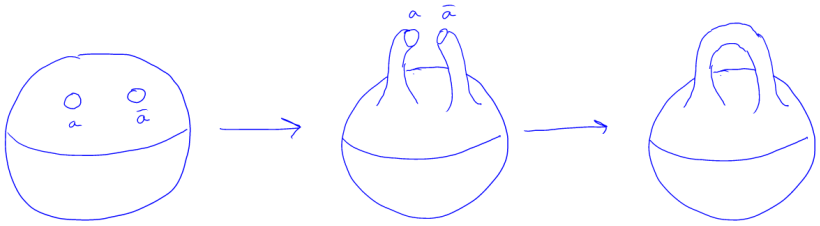


Fig. 8.15 Surgering the twice punctured sphere into a torus. This is the gluing axiom in action.

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

as we have already discussed. The general rule of surgery is that two punctures can be sewed together when they have opposing particle types (i.e., a particle and its antiparticle). 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 S^1. Thus we are sewing together the 2-surface (S^1-puncture $\times$ S^1-time) with another 2-surface (S^1-puncture $\times$ S^1-time), and the inner product between the two wave-functions on these two-surfaces assures that the quantum number on these two punctures are conjugate to each other.

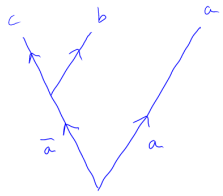


Fig. 8.16 Three particles that fuse to the identity

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. 8.16. 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. 8.17. 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. 8.18 we sew together two pants to obtain a two handled torus.

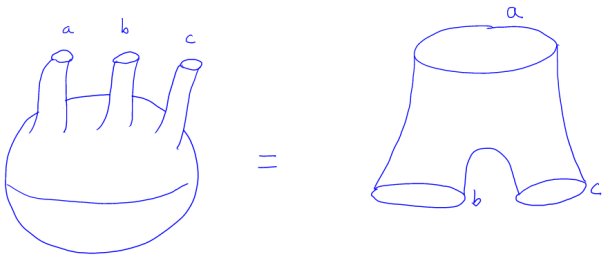


Fig. 8.17 A three-times punctured sphere is known as a "pants" diagram.

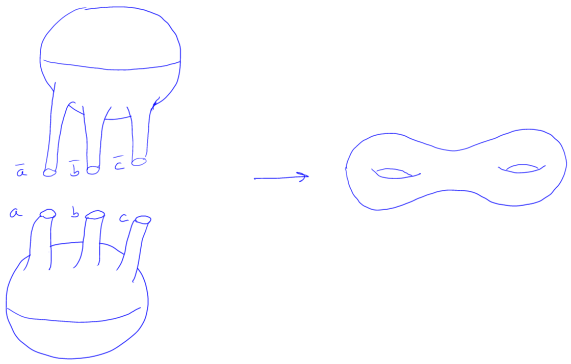


Fig. 8.18 Sewing together two pants diagrams to form a two-handled torus.

To find the ground state degeneracy of the two handled torus,

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

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 $N_{\bar{a}\bar{b}}^{c}$ or N_{abc}). Essentially we are just looking at the number of ways we can assign labels to the punctures when we glue the objects together[8]. Looking at the fusion diagram 8.16 we then have

$$\text{Dim } V(\text{Two handled Torus}) = \sum_{abc} N_{bc}^{\bar{a}} N_{\bar{b}\bar{c}}^{a} = \sum_{abc} N_{abc}^{2}$$

[8] 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.

$$\begin{aligned} I \times I \times I &\to I \\ I \times \tau \times \tau &\to I \\ \tau \times I \times \tau &\to I \\ \tau \times \tau \times I &\to I \\ \tau \times \tau \times \tau &\to I \end{aligned}$$

Here there are several things to note about the notation. First, 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 $\tau \times \tau = I + \tau$ then to indicate that we mean the two τ's fuse in the τ channel only, I will write $\tau \times \tau \to \tau$.

At any rate, there are five possible labelings of the punctures that allow overall fusion to the identity. These are matched together on both

top and bottom of the diagram on the left of Fig. 8.18 and we conclude that in the Fibonacci theory we have

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

8.5 Change of Basis and F-symbols

As mentioned in Fig. 8.14, 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[9] called F. See figure 8.19.

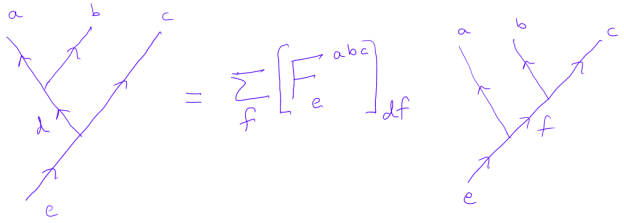


Fig. 8.19 The F-matrix makes a change of basis between the two differnt ways of describing the space spanned by the fusion of three anyon charges a, b, and c when they all fuse to e. For fixed a, b, c and e, the matrix F is unitary in its subscripts d, f. Note: The pictures here are transformations on kets (branches of fusion trees pointing upwards). An F-transformation on bras (reflecting diagrams around the horizontal axis) would require complex complex conjugatation of F.

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.)

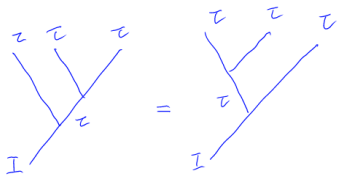


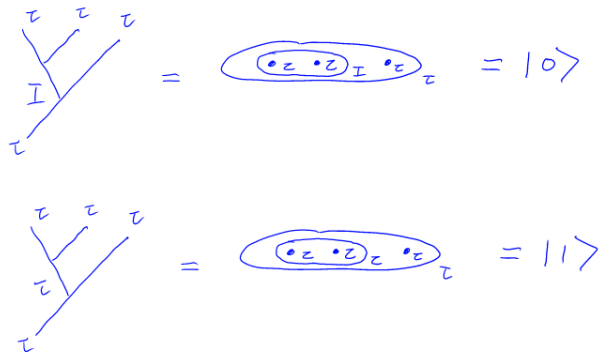
Fig. 8.20 There is only one state in the Hilbert space of three Fibonacci anyons fusing to the identity. Thus it does not matter if you fuse the left two first or the right two first, you are describing the same state. In fact it is possible to define these two states to differ by a phase. However, this is a gauge choice and it is convenient to make gauge choices to get rid of such phases!

Example: Fibonacci Anyons

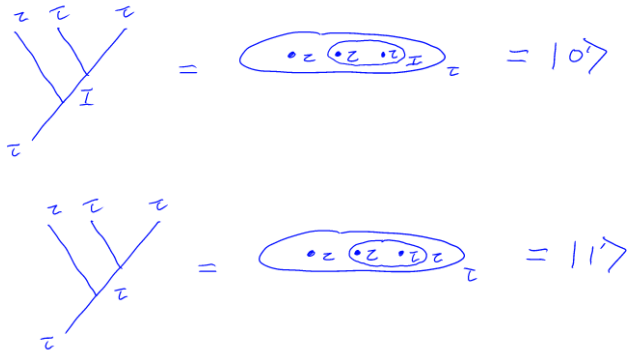
Again we turn to the example of Fibonacci anyons for clarification. We imagine fusing together three τ particles. As shown in Fig. 8.10, there is a single state $|N\rangle$ in which the three fuse to the identity I. It should not matter if we choose to fuse the leftmost two anyons first, or the rightmost two. In either case there is only one possible state for the outcome. We can thus draw the simple identity shown in Fig. 8.20 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

[9] In cases where there are fusion multiplicities are greater than one, each vertex gets an additional index which ranges from 1 to its multiplicity. Most generally this means in Fig.! 8.19 on the left there is an additional index $\mu \in 1 \ldots N_{ab}^{d}$ and $\nu \in 1 \ldots N_{dc}^{e}$ whereas on the right, there is an additional index $\xi \in 1 \ldots N_{bc}^{f}$ and $\eta \in 1 \ldots 1 \ldots N_{af}^{e}$. The F matrix will then be a matrix connecting the space (d, μ, ν) with (f, ξ, η).

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. 8.21. 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. 8.21.



Fusing The Two Particles on the Left First



Fusing The Two Particles on the Right First

Fig. 8.21 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. Again note that we are considering kets so the tree branches point upwards.

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

$$\begin{pmatrix} |0\rangle \\ |1\rangle \end{pmatrix} = \begin{pmatrix} F_{00'} & F_{01'} \\ F_{10'} & F_{11'} \end{pmatrix} \begin{pmatrix} |0'\rangle \\ |1'\rangle \end{pmatrix} \tag{8.6}$$

Here F is a two by two matrix, and in the notaion of the F matrix

defined in Fig. 8.19, 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[10]

$$F = \begin{pmatrix} \phi^{-1} & \phi^{-1/2} \\ \phi^{-1/2} & -\phi^{-1} \end{pmatrix} \tag{8.7}$$

where $\phi^{-1} = (\sqrt{5}-1)/2$ (so ϕ is the golden mean). 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. 8.22, this is precisely the same transformation as in Eq. 8.6, but we must view the cluster of two anyons on the left (underlined), which fuse to τ as being a single τ particle.

[10] Again we can redefine kets with different gauge choices and this will insert some phases into this matrix, but the simplest gauge choice gives the matrix as shown.

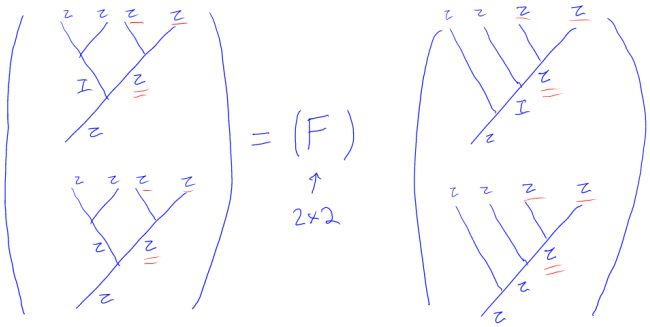


Fig. 8.22 The F-matrices are the same even if one of the anyon charges is made up of a cluster of other anyons. In this particular picture, the cluster of two anyons on the left (underlined in red) has charge τ (double underlined in red). If one were to replace this cluster of two with just a single τ, this would be precisely the same transformation as in Eq. 8.6.

8.5.1 Pentagon

It is possible to describe the same Hilbert space in many ways. For example, with three anyons, as in Fig. 8.14, 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. 8.19, 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. 8.22. 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. 8.23 (the step in the furthest upper left is equilvalent to that shown in Fig. 8.22). 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.

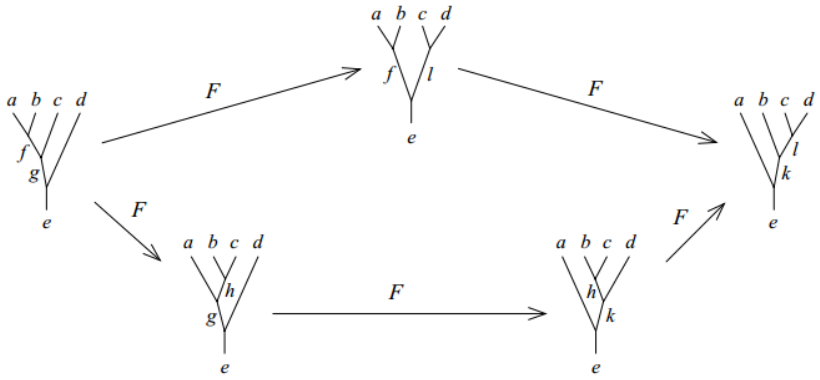


Fig. 8.23 Pentagon Diagram. Figure Stolen from Bonderson's thesis. Each step in the diagram is a new description of the same basis of states via and F-matrix.

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

$$[F_e^{fcd}]_{gl}[F_e^{abl}]_{fk} = \sum_h [F_g^{abc}]_{fh}[F_e^{ahd}]_{gk}[F_k^{bcd}]_{hl} \tag{8.8}$$

where the left hand side represents the top route of the figure and he right hand side represents the bottom route.[11]

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). See exercise ***.

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.

[11] It is very worth working through this to make sure you understand how this equation matches up with the figure!

Further reading

This is some reading.

Exercises

Exercise 8.1 Fibonacci Pentagon

In a TQFT (indeed, in any tensor category), a change of basis is described by the F-matrix as shown in the figure[12].

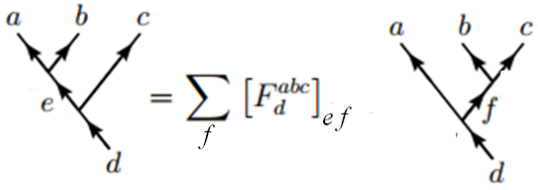


Fig. 8.24 F-matrix

Consistency of F-matrices is enforced by the pentagon equation (Fig. 8.23 and Eq. 8.8).

In the Fibonacci anyon model, there are two particle types which are usually called I and τ. The only nontrivial fusion rule is $\tau \times \tau = I + \tau$. With these fusion rules, the F matrix is completely fixed up to a gauge freedom (corresponding to adding a phase to some of the kets). If we choose all elements of the F matrix to be real, then the F matrix is completely determined by the pentagon. Using the pentagon equation determine the F-matrix. (To get you started, note that the variables a,b,c,d,e,f, g,h can only take values I and τ. You only need to consider the cases where a,b,c,d are all τ).

If you are stuck as to how to start, part of the calculation is given in the Nayak, et al, Rev Mod Phys article (see the reference list)

Exercise 8.2 Fusion and Ground State Degeneracy To determine the ground state degeracy of a 2-manifold in a 2+1 dimensional TQFT one can cut the manifold into pieces and sew back together. One can think of the open "edges" or connecting tube-ends as each having a label given by one of the particle types (i.e., one of the anyons) of the theory. Really we are labeling each edge with a basis element of a possible Hilbert space. The labels on two tubes that have been connected together must match (label a on one tube fits into label $\bar{a}$ on another tube.) To calculate the ground state degeneracy we must keep track of all possible ways that these assembled tubes could have been labeled. For example, when we assemble a torus as in Fig. 8.25, we must match the quantum number on one open end to the (opposite) quantum number on the opposite end. The ground state degeneracy is then just the number of different possible labels, or equivalently the number of different particle types.

For more complicated 2-d manifolds, we can decompose the manifold into so-called pants diagrams that look like Fig. 8.26. when we sew together pants diagrams, we should include a factor of the fusions multiplicity N_{ab}^c for each pants which has its three tube edges labeled with a, b and $\bar{c}$.

(a) Write a general formula for the ground state degeneracy of an M-handled torus in terms of the N matrices.

[12]Strictly speaking when there are fusion multiplicities, $N_{bc}^a > 1$, then one also needs an additional index at each vertex.

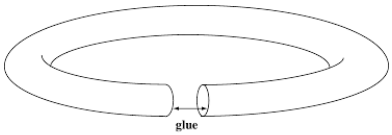


Fig. 8.25 Gluing together a torus

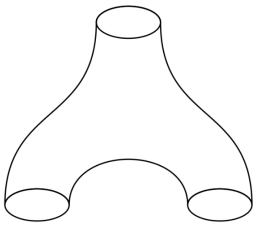


Fig. 8.26 A pants diagram

(b) For the fibonacci anyon model, find the ground state degeneracy of a 4-handled torus.

Exercise 8.3 Quantum Dimension

Let N_{ab}^c be the fusion multiplicity matrices of a TQFT

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

meaning that N_{ab}^c is the number of distinct ways that a and b can fuse to c. (In many, or even most, theories of interest all N's are either 0 or 1).

The quantum dimension d_a of a particle a is defined as the largest eigenvalue of the matrix $[N_a]_b^c$ where this is now thought of as a two dimensional matrix with a fixed and b, c the indices.

Show that

$$d_a d_b = \sum_c N_{ab}^c \, d_c$$

Try to prove this without invoking the Verlinde formula which we run into later in chapter ***.

Exercise 8.4 Ising F-matrix

[Hard] As discussed in the earlier problem, "Ising Anyons and Majorana Fermions", one can express Ising anyons in terms of Majorana fermions. As discussed there we can choose any two majoranas and construct a fermion operator

$$c_{12}^\dagger = \frac{1}{2}(\gamma_1 + i\gamma_2)$$

then the corresponding fermion orbital can be either filled or empty. We might write this as $|0_{12}\rangle = c_{12}|1_{12}\rangle$ and $|1_{12}\rangle = c_{12}^\dagger|0_{12}\rangle$. The subscript 12 here meaning that we have made the orbital out of majoranas number 1 and 2. Note however, that we have to be careful that $|0_{12}\rangle = e^{i\phi}|1_{21}\rangle$ where ϕ is a gauge choice which is arbitrary (think about this if it is not obvious already).

Let us consider a system of 4 majoranas, $\gamma_1, \gamma_2, \gamma_3, \gamma_4$. Consider the basis of states

$$\begin{aligned} |a\rangle &= |0_{12}0_{34}\rangle \\ |b\rangle &= |0_{12}1_{34}\rangle \\ |c\rangle &= |1_{12}0_{34}\rangle \\ |d\rangle &= |1_{12}1_{34}\rangle \end{aligned}$$

rewrite these states in terms of basis of states

$$\begin{aligned} |a'\rangle &= |0_{41}0_{23}\rangle \\ |b'\rangle &= |0_{41}1_{23}\rangle \\ |c'\rangle &= |1_{41}0_{23}\rangle \\ |d'\rangle &= |1_{41}1_{23}\rangle \end{aligned}$$

Hence determine the F-matrix for Ising anyons. Be cautious about fermionic anticommutations: $c_x^\dagger c_y^\dagger = -c_y^\dagger c_x^\dagger$ so if we define $|1_x 1_y\rangle = c_x^\dagger c_y^\dagger |0_x 0_y\rangle$ with the convention that $|0_x 0_y\rangle = |0_y 0_x\rangle$ then we will have $|1_x 1_y\rangle = -|1_y 1_x\rangle$. Note also that you have to make a gauge choice of some phases (analogous to the mentioned gauge choice above). You can choose F to be always real.

9 Planar Diagrams

As we discussed in chapter 7, one of our objectives is to come up with some diagrammatic rules (somewhat analogous to those of the Kauffman invariant) which will allow us to evaluate any diagram of world-lines (i.e, a labeled link) and get an output which is a complex number. Having described the idea of the F-matrix we can begin to construct these rules. In this chapter we will focus only on planar diagrams — i.e., we do not allow lines to cross over each other forming braids. We can think of this as being a definition of a TQFT in 2+1 dimensions — particles moving in 1+1 dimension. Since there are no over and under-crossings the only nontrivial possibility is that particles come together to fuse, or split apart. An example of a planar fusion diagram is shown in Fig. 9.1. It is convenient to draw diagrams so that no lines are drawn exactly horizontally. The reader should be cautioned that there are several different normalizations of diagrams (two in particular). These are convenient in different contexts and one should be warned not to confuse them. We will start with a more "physics" oriented normalization in section 9.1, but the we switch to a more topologically oriented normalization in section ***.

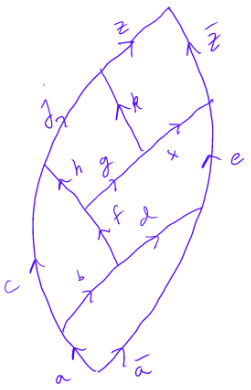


Fig. 9.1 A planar fusion diagram

9.1 Physics Normalization

To make sense of these diagrams we should think in terms of the anyon Hilbert spaces defined in the chapter 8. Each fusion tree we write (starting with the vacuum at the bottom) corresponds to a particular state vector of the Hilbert space. Two fusions trees with the same branching structure are orthogonal if any of their labels differ. For example, we have the orthogonormality shown in Fig. 9.2[1]

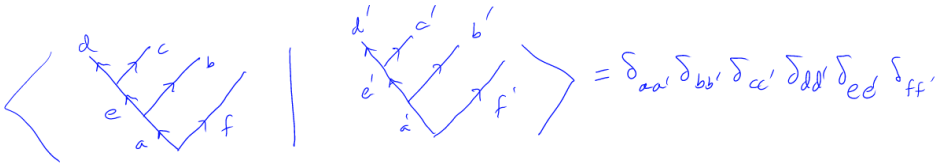


Fig. 9.2 Orthogonormality of trees with the same branching structure but different indices. Note that the normalization of diagrams will be changed in section ** below

Note that the normalization of these diagrams will be changed in section *** below. Typically, we draw a bra by turning over the diagram

[1] To be precise, in cases where there are fusion multiplicities greater than one, each vertex get an adiditional index (See margin note 9 from chapter 8). In this case, the diagrams are also orthonormal in these indices as well. For simplicity of notation we will not write these vertex indices.

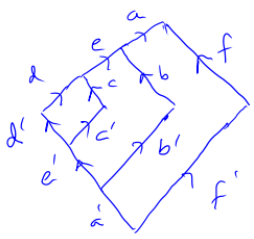


Fig. 9.3 Expressing the inner product of Fig. 9.2 as a single diagram.

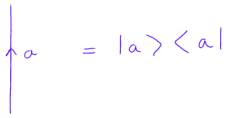


Fig. 9.4 A labeled straight line is just an identity operator for the par-

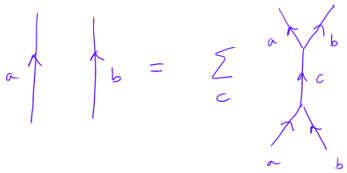


Fig. 9.5 Insertion of a complete set of states.

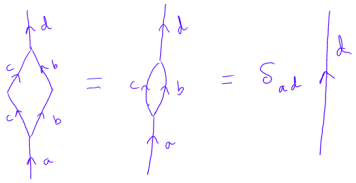


Fig. 9.6 This identity is implied by the orthogonality of trees with the same branching structure, but it is also a result of the locality principle.

(reflecting it around a horizontal axis) so the diagram in Fig. 9.2 could be re-expressed as the diagram shown in Fig. 9.3.

Our convention is that when we draw a diagram with world-lines that end pointing upwards (like the right ket of Fig. 9.2), we mean that these particles are kets. If world lines end pointing downwards, we mean them to be bras (We have used this convention in the upper half of Fig. 9.3, but in the left of Fig. 9.2 we do not use this convention). Many diagrams will have lines that point both up and down (like in Fig. 9.8 below or Fig. 8.19 above). We can think of these pictures as being just a piece of a larger diagram. Equivalently we mean that the diagram has some particles that live in the vector space of kets and some in the dual (bra) space. Such diagrams can be interpreted as operators that take as input the lines coming in from the bottom and give as output the lines going out the top.

A particularly important diagram is given by a simple labeled straight line, which should be interpreted as the identity operator for the particle label type as shown in Fig. 9.4. One can insert a complete set set of states into straight lines to obtain the result shown in Fig. 9.5. We should see this as a ket on top of a bra $\sum_x |x\rangle\langle x|$ which is our usual idea of a complete set. Another very useful diagram is given by Fig. 9.6. This is implied by the orthogonality of tree states, but it also is a manifestation of the locality principle shown in Fig. 8.7.

A few simple principles are worth noting.

(1) one is free to add or remove lines from a diagram if they are labeled with the identity or vacuum (I). So for example, we can write the diagram on the left of Fig. 9.8 as a tree as shown on the right of the figure.

(2) Starting from the vacuum, a particle can only be pair created with its antiparticle, and similarly two particles can only pair annihilate to the vacuum if they are antiparticles. So, for example, this tells us that the diagram in Fig. 9.3 can only be nonzero if $a = f$ and $a' = f'$.

(3) A line must maintain its quantum number unless it fuses with another line, or splits. So for example, this tells us immediately in Fig. 9.3 that $d = d'$ and $c = c'$ and $b = b'$ and $f = f'$.

(4) One can use F-moves to change the structure of fusion trees in order to determine their values. For example, in Fig. 9.7, the diagram on the left is turned into the one on the right using an F-move, and then the diagram on the right is easily evaluated since both the upper and lower halfs have the same tree structure, as in Fig. 9.2 and Fig. 9.3.

In fact, with these principles (and given an F-matrix as input information) it is possible to fully evaluate any planar diagram into a complex number. It should be emphasized, however, that with these rules diagrams are not invariant under all smooth deformations, as we shall now see.

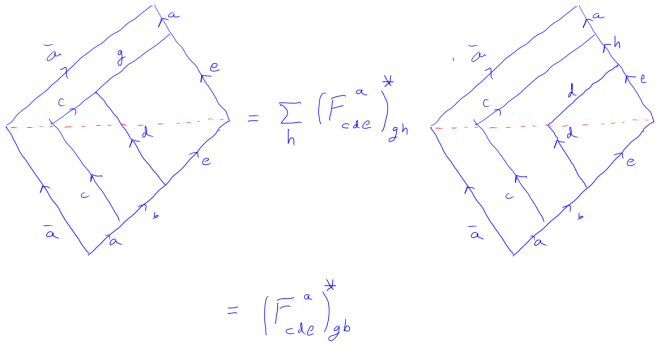


Fig. 9.7 The diagram on the left is evaluated by applying an F-move to the upper half of the diagram. If the diagram is cut in half along the dotted line and the upper half is flipped over, it is transformed exactly as shown in Fig. 8.19. The flipping over of this part of the diagram turns the ket into a bra, and correspondingly the F matrix has been complex conjugated. The diagram on the right is evaluated to delta functions as in Figs. 9.2 and 9.3

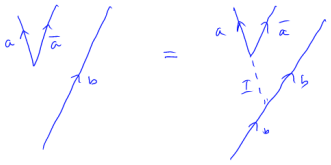


Fig. 9.8 Lines labeled with the identity, or vacuum, can always be added or removed from diagrams.

9.2 Switching to Planar Isotopy Invariant Normalization

We would like to have a diagrammatic planar calculus for our evaluation of diagrams. We would like to have rules such that we can "straighten" out lines as suggested in Fig. 9.9.

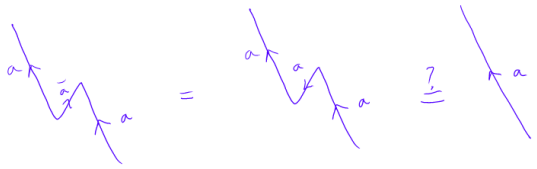


Fig. 9.9 We would like to build diagrammatic rule so that lines that have up-down wiggles can be straightened out.

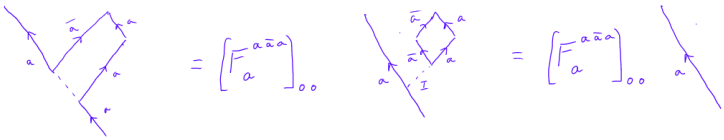


Fig. 9.10 Removing a wiggle from a diagram incurs a factor of $[F_a^{a\bar{a}a}]_{00}$.

However, we can straighten a wiggle using an F-move, as shown in Fig. 9.10 and we discover that it incurs a factor of $[F_a^{a\bar{a}a}]_{00}$ where the

subscript 0 means the identity particle. It is conventional to define[2]

$$d_a = \frac{1}{|[F_a^{a\bar{a}a}]_{00}|} \tag{9.1}$$

which will end up being the quantum dimension of the particle (See chapter 12).

We would like to have diagrammatic rules such that we can straighten wiggles like those in Fig. 9.9 freely. To arrange this, let us define some new diagrammatic rules, which attaches a factor of d_a to each loop.

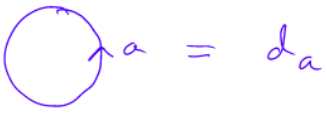


Fig. 9.11 Isotopy invariant diagrammatic rule 1

Assigning a closed loop "world-line" a factor of d_a is familiar from our exploration of the Kauffman invariant in chapter 2. Similar as what we saw in section 2.3 this means that bras and kets will not have conventional normalization $\langle x|x\rangle = 1$. However, paying this price allows us to have diagrams where wiggly lines can be straightened as in Fig. 9.9. Attempting the calculation shown in Fig. 9.10 with the new normalization, we see that the factor of d_a for obtaining the closed loop cancels the factor of $[F_a^{a\bar{a}a}]_{00}$. One might be worried that d_a is actually defined (in Eq. 9.1) as the inverse of the *absolute value* of this F factor. However, it turns out that in most cases, one can absorb any phase into the definition of the creation or annihilation vertex as explained in Fig. 9.12. As such, one usually chooses these phases so that $[F_a^{a\bar{a}a}]_{00}$ is real and positive. We will assume that we always make this gauge choice, so that wiggles like that in Fig. 9.9 can always be straightened freely.

There is, however, one possibile situation that ruins this entire argument! In the case where a particle is its own antiparticle, $a = \bar{a}$, then there is not enough gauge freedom to remove the phase that one can incur when straightening out the wiggle[3]. In this case our straightening rule is as shown in Fig. 9.13: removing a wiggle incurs a factor of κ_a, the so called Frobenius-Schur indicator. This quantity is a (gauge-invariant) fundamental property of the particle in the theory and must always be either $+1$ or -1. Particles with $\kappa_a = +1$ have full isotopy invariance — and deformation of a curve is acceptable. Particles with $\kappa_a = -1$ do not have full isotopy invariance and one must keep track of how many wiggles are introduced! This may seem strange, but, as we point out in the appendix of this chapter, this physics is essentially already present with angular momentum addition.

Fig. 9.12 The vertex $|a\bar{a}\rangle$ (shown on the left) and the vertex $|\bar{a}a\rangle$ (on right) can be assigned different phases as a gauge choice. In Fig. 9.10 the leftmost figure includes $|a\bar{a}\rangle$ and $\langle\bar{a}a|$, Whereas the phases cancel in the loop formed in the middle picture of Fig. 9.10. Thus choosing gauge correctly we can generally make $[F_a^{a\bar{a}a}]_{00}$ real and positive, with the exception of cases where $a = \bar{a}$.

[2]For some theories with nontrivial Frobenius-Schur indicator, it is convenient to define d_a to be negative. However, most generally we want to define it positive. See section *** on Frobenius-Schur below.

[3]Note that the creation kets in Fig. 9.12 are the same if $a = \bar{a}$ so one cannot insert a phase on one but not the other!

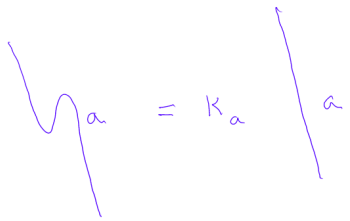


Fig. 9.13 Removing a wiggle from a self-dual particle $a = \bar{a}$ incurs a factor of the Frobenius-Schur indicator, κ_a. This is a gauge invariant quantity, but must be either $+1$ or -1 for any self-dual particle. This is a fundamental property of the particular particle type. For any non-self-dual particle we can always choose a gauge so that straightening a wiggle does not incur any phase.

We note that in some theories one may be able to remove the non-trivial Frobenius-Schur indicator (and obtain diagrams with full isotopy invariance) by choosing a definition of d_a to be negative rather than positive. We have seen examples of this when we considered the Kauffman invariant — it is often allowable to have $d < 0$. However, having a negative value for d is inconvenient in other ways. If we want to think of a single loop as being $d = \bigcirc = \langle \cup | \cup \rangle$, or a pair creation ket on the lower half joined with its conjugate pair annhiliation bra on the upper half, we have the unusual problem that the state

FILL IN MORE HERE...

10 Braiding and Twisting

10.1 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

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

In our diagrammatic notation, we have twists factors defined by Fig. 10.1. 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.

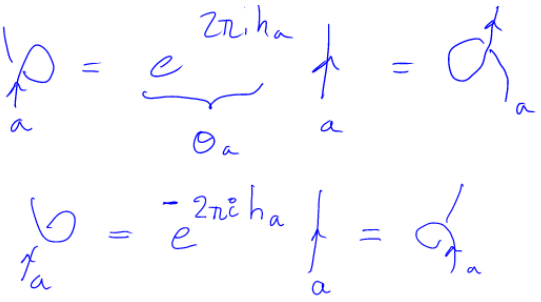


Fig. 10.1 Definition of Twist Factor

Recall we should treat particle world-lines as ribbons, so that a loop can be pulled straight as in Fig. 10.2 to represent a particle twisting around its own axis, as well as giving the phase of exchange for two identical particles (See also Fig. 2.7). 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, $h_I = 0$.

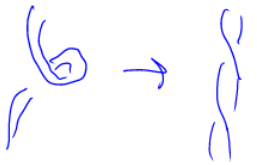


Fig. 10.2 Pulling a ribbon straight

10.2 R-matrix

Consider the possibility of two particles fusing to a third as shown in Fig. 10.3. 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

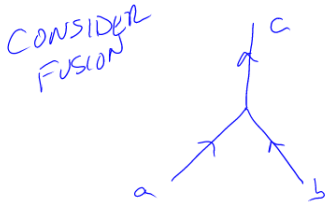


Fig. 10.3 Two particles fusing to a third. For this anyon system $a \times b = c + \ldots$, and c is the particular fusion channel that has occured in this diagram.

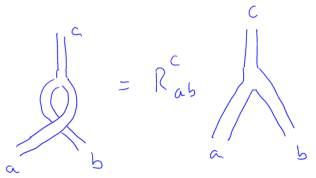


Fig. 10.4 Definition of R-matrix. Here we drop the arrows for convenience of notation and we draw the particle world lines as ribbons to show that no additional self-twists are incurred by the particles.

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. 10.4. 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. Note that braiding anything with the identity particle should be trivial, $R^a_{Ia} = R^a_{aI} = 1$.

Taken together with the F-matrices, the R-matrix allows us to calculate the physical result of any braid, as we shall see below.

To see the relationship between braiding and twisting, consider applying the R matrix twice to make a double twist as in Fig. 10.5. 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.

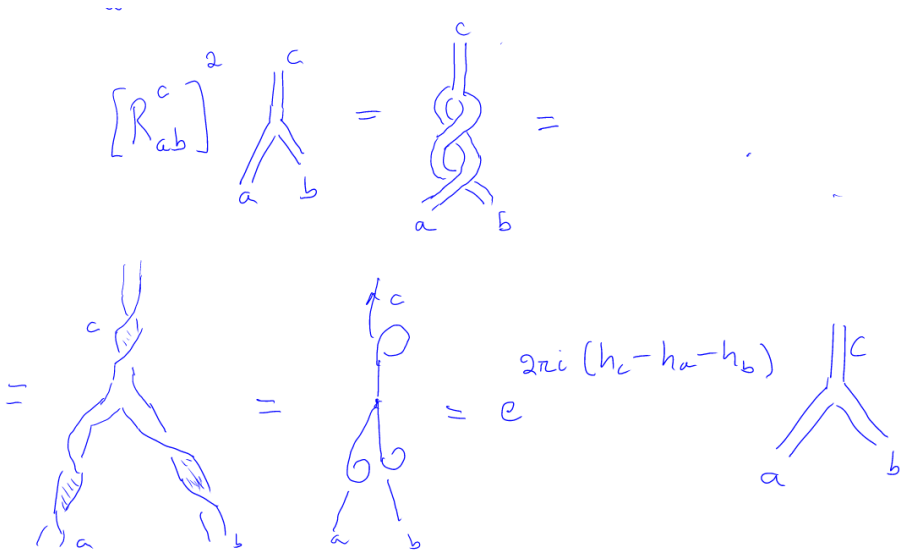


Fig. 10.5 Relation of R-matrix to twist factors.

We can generally write this relationship as

$$[R^c_{ab}]^2 = e^{2\pi i(h_c - h_a - h_b)} = \theta_c/(\theta_a\theta_b) \tag{10.1}$$

Example: Fibonacci Anyons

In the Fibonacci theory, two τ particles can fuse to either τ or I. Applying the above relationship, Eq. 10.1, we have

$$[R^\tau_{\tau\tau}]^2 = e^{2\pi i(h_\tau - h_\tau - h_\tau)} = e^{-2\pi i h_\tau} \tag{10.2}$$

$$[R^I_{\tau\tau}]^2 = e^{2\pi i(h_I - h_\tau - h_\tau)} = e^{-4\pi i h_\tau} \tag{10.3}$$

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. 10.6. Now consider

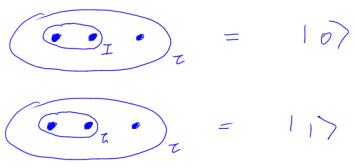


Fig. 10.6 The two states of three τ particles fusing to τ. Unmarked dots are τ particles.

braiding the two leftmost particles around each other.

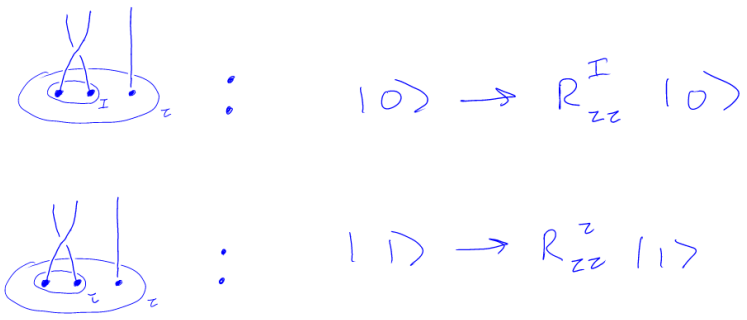


Fig. 10.7 Braiding the two left particles in this basis gives a phase dependent on the fusion channel of the two particles.

The result of this braiding 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.

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

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. 10.8? 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. If we want, we can then use the F-matrix to transform back to the original basis. To see how this works, Recall that we can use the F matrix to write (See Eq. 8.6)

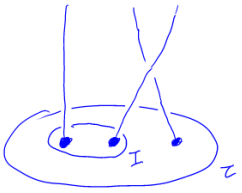


Fig. 10.8 How does one evaluate this braid? One applies F-first, then R as shown in the next two figures!

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

or in diagrams (see Fig. 10.9).

Fig. 10.9 The F-matrix relation in diagram form. See Eq. 8.6

On the right hand side of Fig. 10.9 (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.

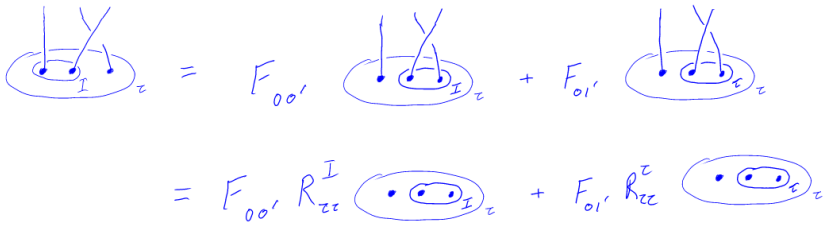


Fig. 10.10 To braid particles, switch basis using F until we know the fusion channel of the two particles we want to braid, and then we can apply the R-matrix.

To describe this in equations, we can write the operator that braids the rightmost two particles as R_{23} and then we have

$$\begin{aligned}
R_{23}|0\rangle &= R_{23}(F_{00'}|0'\rangle + F_{01'}|1'\rangle) \\
&= F_{00'}R_{23}|0'\rangle + F_{01'}R_{23}|1'\rangle \\
&= F_{00'}R^{I}_{\tau\tau}|0'\rangle + F_{01'}R^{\tau}_{\tau\tau}|1'\rangle \qquad (10.4)\\
&= F_{00'}R^{I}_{\tau\tau}\left([F^{-1}]_{0'0}|0\rangle + [F^{-1}]_{0'1}|1\rangle\right) \qquad (10.5)\\
&\quad + F_{01'}R^{\tau}_{\tau\tau}\left([F^{-1}]_{1'0}|0\rangle + [F^{-1}]_{1'1}|1\rangle\right) \\
&= \left(F_{00'}R^{I}_{\tau\tau}[F^{-1}]_{0'0} + F_{01'}R^{\tau}_{\tau\tau}[F^{-1}]_{1'0}\right)|0\rangle \\
&\quad + \left(F_{00'}R^{I}_{\tau\tau}[F^{-1}]_{0'1} + F_{01'}R^{\tau}_{\tau\tau}[F^{-1}]_{1'1}\right)|1\rangle
\end{aligned}$$

Where between Eq. 10.4 and 10.5 we have used the inverse F transform to put the result back in the original $|0\rangle$ and $|1\rangle$ basis.[1].

This general principle allows us to evaluate any braiding of particles. We can always convert to a basis where the fusion channel of the two particles to be 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.

[1]For this particular case (using Eq. 8.7 for the F-matrix) the matrix F and F^{-1} happen to be the same matrix (however we write out the inverse explicitly for clarity!)

10.3 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. 10.11. In equations this can be expressed as

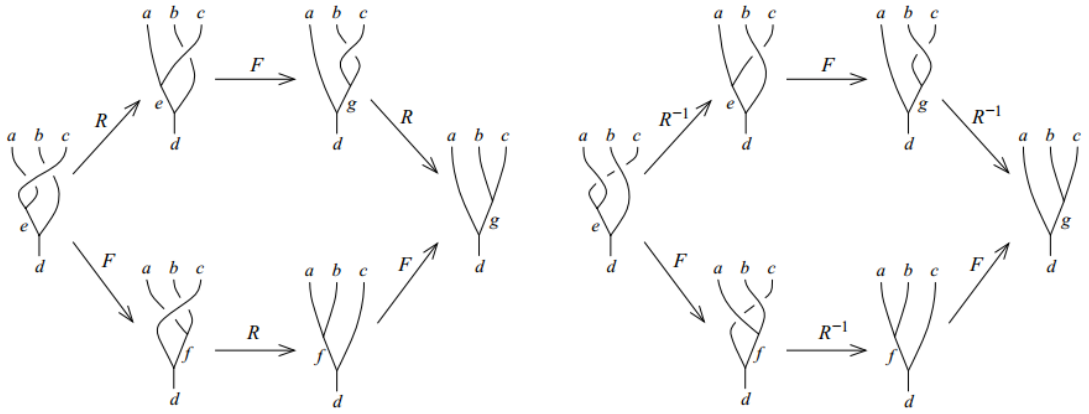


Fig. 10.11 The hexagon equations in graphical form. (nice picture stolen from Bonderson thesis)

$$
\begin{aligned}
R^{ca}_{e}[F^{acb}_{d}]_{eg}R^{cb}_{g} &= \sum_{f}[F^{cab}_{d}]_{ef}R^{cf}_{d}[F^{abc}_{d}]_{fg} \\
\left[R^{ca}_{e}\right]^{-1}\left[F^{acb}_{d}\right]_{eg}\left[R^{cb}_{g}\right]^{-1} &= \sum_{f}[F^{cab}_{d}]_{ef}[R^{cf}_{d}]^{-1}[F^{abc}_{d}]_{fg}
\end{aligned}
$$

The top equation is the left diagram whereas the lower equation is the right diagram in Fig. 10.11. 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 (See exercise 10.1).

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

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

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

10.4 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 the pentagon up to a gauge freedom 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[2]. 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 five or six different particle types[3,4]. 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.

[2]But not published by him! See for example, "On fusion categories", Annals of Mathematics, Pages 581-642 from Volume 162 (2005), by Pavel Etingof, Dmitri Nikshych, Viktor Ostrik.

[3]See for example E. Rowell, R. Stong, and Z. Wang, On Classification of Modular Tensor Categories, Comm. Math. Phys. 292 (2009),p343. arXiv:0712.1377

[4]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!

Further reading

This is some reading

Exercises

Exercise 10.1 Fibonacci Hexagon In any TQFT or "braided" (including modular) tensor category (think of all of these as just anyon theories! don't worry about the fine distinctions), a braiding is defined by an R-matrix as shown in the figure 10.4. Once F matrices are defined for a TQFT, consistency of the R-matrix is enforced by the so-called hexagon equations as shown in the figure diagramatically by Fig. 10.11.

For the Fibonacci anyon theory, once the F matrix is fixed as in Eq. 8.7, the R matrices are defined up to complex conjugation (i.e., there is a right and left handed Fibonacci anyon theory — both are consistent). Derive these R matrices.

11

Kauffman Anyons

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. 2.3. 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. 11.1. 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.

Fig. 11.1 The loop rule for the Kauffman invariant and 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[1] 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. 11.2 The fact that two 1-particles can fuse to the vacuum tells us immediately that 1 is its own antiparticle.

$$1 = \bar{1}$$

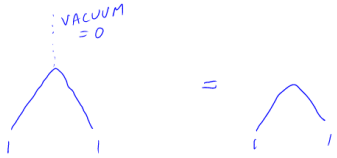


Fig. 11.2 Fusing two 1-particles to the vacuum

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. 11.3. 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 (see section 8.6). 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.

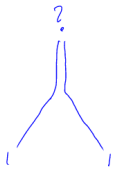


Fig. 11.3 Attempting to Fuse two 1-particles to something different from the vacuum

11.1 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 as shown in Fig. 11.4. 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

Fig. 11.4 The projector of two strings to the vacuum P_0.

[1] I realize here this is confusing to have 1 not be the identity, but this is the usual notation! It is (not coincidentally!) similar to spins where spin 0 is the identity (no spin), and spin 1 is

operator to two strings coming in the bottom of the second operator. As shown in Fig. 11.5, 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.

Fig. 11.5 Checking that $P_0^2 = P_0$.

Fig. 11.6 The projector of two strings to the nontrivial particle $P_2 = I - P_0$.

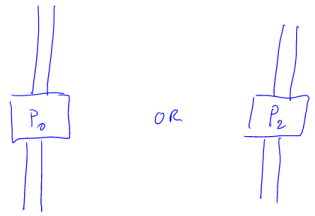


Fig. 11.7 Drawing the two possible fusion channels of two strings as a box labeled P_0 or P_2

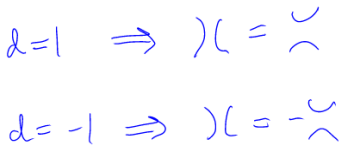


Fig. 11.8 Two cases where the Kauffman invariant rules become very simple. If you have not convinced yourself of these rules, try to do so! (See exercise 2.2). Note that $d = 1$ occurs for bosons or fermions and $d = -1$ occurs for semions.

We now consider the possibility that two strings near each other can fuse to something else (not the identity). Let a projector onto this particle type be called P_2, i.e., we define $P_2 = I - P_0$ where I is the identity operator, or just two parallel strings. Diagrammatically we have Fig. 11.6 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 show $P_2 P_0 = 0$.

Often it is convenient to draw these projection operators as a labeled box, as shown in Fig. 11.7. Sometimes we simply draw a single line with a label, 0 or 2 respectively.

Abelian Case

In the case where $d = \pm 1$ it is easy to prove (see Exercise 2.2) that two horizontal lines equals ± 1 times two vertical lines as shown in Fig. 11.8. In this case, notice that the projector $P_2 = 0$ (the two terms in the projector in Fig. 11.6 are equal with opposite signs). 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. 11.2. Thus the entire fusion rules of these theories are

$$1 \times 1 = 0$$

where again 0 is the identity or vacuum.

Two Strands in the 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. 11.9. We can write the fusion rule as

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

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 $P_0 + P_2 = I$, which means these two particle types form a complete set.

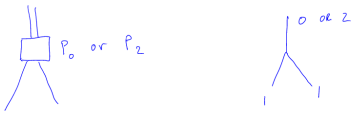
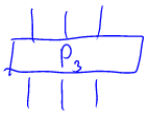


Fig. 11.9 Possible fusions of two strands

Three Strands in the General Case

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. 11.10.



General Form:

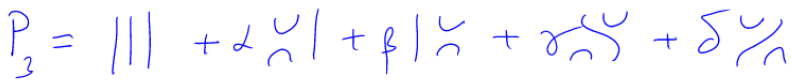
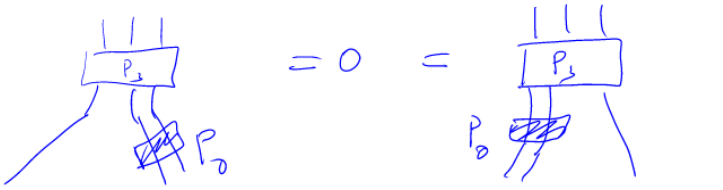


Fig. 11.10 A three-leg Jones-Wenzl projector. The most general possible form of P_3 is written in the lower line.

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 this means we must insist on relations like Fig. 11.11.



or

Equivalently

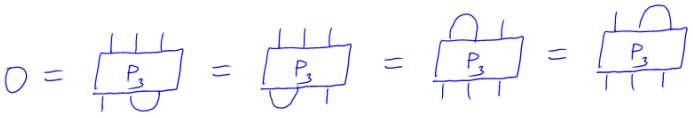


Fig. 11.11 Insisting that 0×1 does not give 3

However, we should alllow fusions of the form $1 \times 2 = 3$ as shown in Fig. 11.12 Enforcing the condition in Fig. 11.11, along with $P_3^2 = P_3$ gives the form of P_3 shown in Fig. 11.13. (See Exercise 11.1).

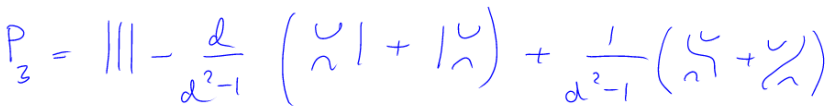


Fig. 11.13 Form of the P_3 projector in terms of the parameter d.

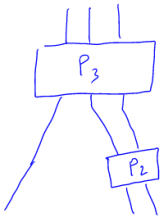


Fig. 11.12 We allow $1 \times 2 = 3$

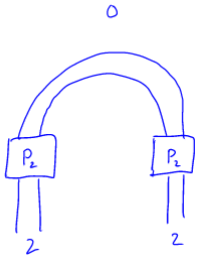


Fig. 11.14 $2 \times 2 = 0$.

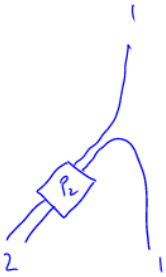


Fig. 11.15 $2 \times 1 = 1$. We recognize this as the fusion $1 \times 1 = 2$ from Fig. 11.9 just turned on its sid.e

Ising Anyons

Consider the case where $d = \sqrt{2}$. Here it is possible to show that P_3 vanishes when evaluated in any diagram (See exercise 11.1). 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. 11.14 and $2 \times 1 = 1$ as shown in Fig. 11.15. (Note that showing $2 \notin 2 \times 2$ requires another explicit calculation, not shown here! See exercise 11.1) We thus have the full set of nontrivial fusion rules

$$\begin{aligned} 1 \times 1 &= 0 + 2 \\ 2 \times 2 &= 0 \\ 1 \times 2 &= 1 \end{aligned}$$

which we recognize as Ising fusion rules (see 8.2.2) where $1 = \sigma$ and $2 = \psi$.

Note: It is not coincidence that the quantum dimension of the non-abelian particle (the 1-particle, or σ) is $\sqrt{2}$ (see section 8.2.2), 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.

11.2 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

$$F = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$

where the columns and rows of this matrix represent quantum numbers 0 and 2. By this F-matrix we mean the diagram Fig. 11.16

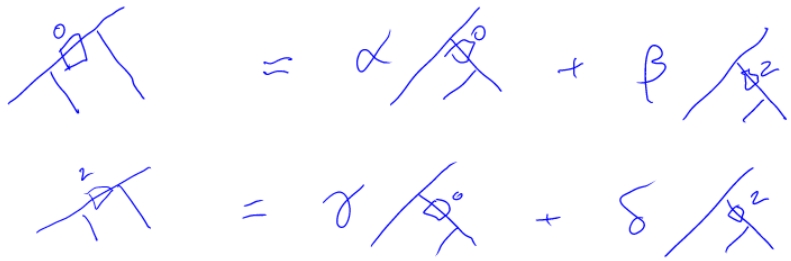


Fig. 11.16 An F-move. The boxes are either P_0 or P_2 projectors. Unmarked lines are all single lines or "1".

We now recall that graphically a P_0 projector is $1/d$ times a simple turnaround (see Fig. 11.4), whereas a P_2 projector (Fig. 11.6) is the sum of two terms, the first being just two strands going parallel (i.e,, the identity) and the second being $-P_0$. We can then write graphically Fig. 11.17

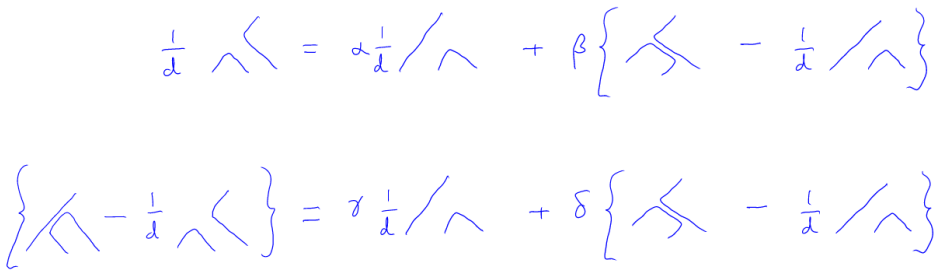


Fig. 11.17 Plugging in the form of the projectors into Fig. 11.16.

We then match up terms on the right and left of these graphical equations. 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 brackets on the left the same as the first term in brackets on the right, so we have $\delta = -1/d$. Then among the remaining terms, the first term in brackets on the left, the first term on the right, and the second 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 we obtaini the full form of the F matrix

$$F = \begin{pmatrix} 1/d & 1/d \\ d - 1/d & -1/d \end{pmatrix}$$

11.3 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. 2.3. As shown in Fig. 2.6, comparing to Fig. 10.1 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. 11.18

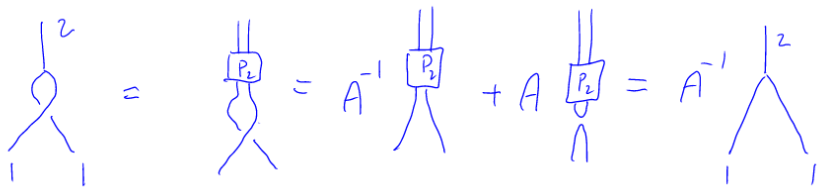


Fig. 11.18 Evaluation of $R^2_{11} = A^{-1}$.

Here the term with the coefficient A vanishes because of the orthogonality of P_2 and P_0.

Further Reading

- Louis Kauffman, *Knots and Physics*, World Scientific, (2001), 3ed.
- L. H. Kauffman and S. L. Lins, *Temperley-Lieb Recoupling Theory and Invariants of 3-Manifolds*, Annals of Mathematics Studies, no 134, Princeton University Press (1994).

Exercises

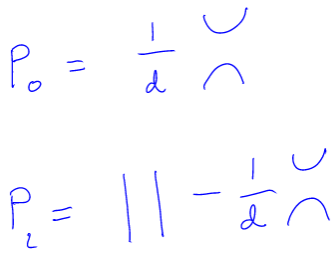


Fig. 11.19 Jones-Wentzl projectors for two strands

Exercise 11.1 Jones-Wenzl projectors The Temperly-Lieb algebra is the algebra of strings in two dimension that are not allowed to cross. As with the Kauffman invariant, a closed loop gets a value d. As in the lecture, we can "cable" together two strings to make two types of "particles". Projectors onto these two particles types are given by the so-called Jones-Wenzl projectors shown in Fig. 11.19

(a) Show that the projectors satisfy $P^2 = P$, so their eigenvalues are 0 and 1. Further show that the two projectors are orthogonal $P_0P_2 = P_2P_0 = 0$. (should be easy, we did this in lecture)

(b) Show that for $d = \pm 1$ we have $P_2 = 0$ in the evaluation of any diagram. The result means that in these models there is no new particle which can be described as the fusion of two elementary anyons. Why should this be obvious? Hint: Look back at the the problem called "Abelian Kauffman Anyons".

(c) Using the Jones-Wenzl projectors with the Kauffman braiding rules, and choosing $A = i^{3/4}$ (which corresponds to the anyons of the $\nu = 5/2$ fractional quantum Hall state), show that the P_2 particle is a fermion, (The P_0 particle is a boson, but this should be trivial!). Hint: Show this by calculating the twist factor of the P_2 particle — it is easier, but equivalent to, calculating exchange of two particles.

(d) The three strand Jones-Wenzl projector must be of the form shown in the figure 11.20.

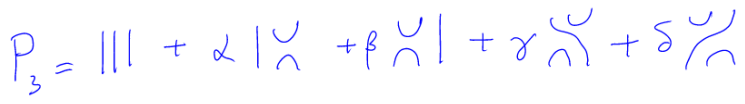


Fig. 11.20 Jones-Wentzl projector for three strands

The coefficients $\alpha, \beta, \gamma, \delta$ are defined by the projector condition $P_3^2 = P_3$ and also by the condition that P_3 is orthogonal to P_0 which is shown in the figure 11.21.

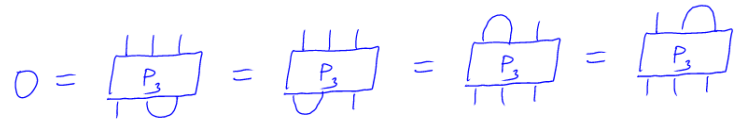


Fig. 11.21 Condition on Jones-Wentzl projector for three strands

Calculate the coefficients $\alpha, \beta, \gamma, \delta$ in P_3.

(e) Choosing $A = i^{3/4}$ again, show that $P_3 = 0$ in the evaluation of any diagram. We can then conclude that in this model there is no new particle that is the fusion of three elementary strands. Hint: think about how the "Abelian Kauffman Anyon" problem was solved.

Diagrammatic Algebra, the S-matrix and the Verlinde Relation

12

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) as the particles move around in the manifold.

We are almost at the point where we have a full diagramatic calculus — which would produce a number as an output given any world-line diagram as input

$$Z(\text{Manifold with particle world lines in it}) \to \mathbb{C}$$

Note that while the 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's thesis (*** See also chapter *** to be added).

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

$$\begin{aligned} \langle \text{Link} \rangle &\equiv \frac{Z(S^3 \text{ with embedded Link})}{Z(S^3)} \\ &= Z(S^2 \times S^1 \text{ with embedded Link}) \end{aligned}$$

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 7).

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.

12.1 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. 12.1[2]. These quantities will turn out to be the

[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

quantum dimensions of the particles, but we have not shown this yet!

Fig. 12.1 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 7.5 to note that we have

$$Z(S^3; a \text{ loop linking } b \text{ loop}) = S_{ab} = S_{ba}$$

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[3]).

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

$$1 = \sum_a |S_{a0}|^2 = |S_{00}|^2 \sum d_a^2$$

where the sum is over all particle types in the theory. We can then write[4]

$$Z(S^3) = S_{00} = 1/\mathcal{D}$$

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.

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 and will be discussed in section ***.

[3]It turns out that any fermion will make a theory non-modular! This is why fermions are a bit difficult to handle!

[4]See footnote 9 in chapter 7 for why S_{00} must be real.

12.2 Quantum Dimensions

Now, we claim that these loop quantites d_a should satisfy the fusion algebra

$$d_a d_b = \sum_c N_{ab}^c \; d_c \tag{12.1}$$

or diagrammatically we have Fig. 12.2

Fig. 12.2 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 (See also Exercise 8.3). However, to prove it is a bit more complicated than this argument, and is given in the appendix to this chapter.

Now, given Eq. 12.1, 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[5] 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!

12.3 Verlinde Algebra

Using the locality principle (or no-transmutation) principle (See Fig. 8.7) we can show that a closed loop of type a around a world line of type x

[5]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.

gives some constant which we call $\tilde{S}_{ax}$ as shown in Fig. 12.3.

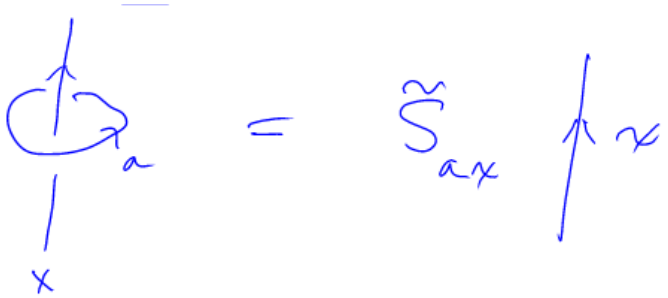


Fig. 12.3 The locality priniciple tells us that the value of a loop around a world line is some number which we call $\tilde{S}_{ax}$

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

$$\tilde{S}_{ax} = \frac{S_{ax}}{S_{0x}} \tag{12.2}$$

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.12.4. This identity is entirely analogous to that of Fig. 12.2, and the rigorous derivation is given in the appendix.

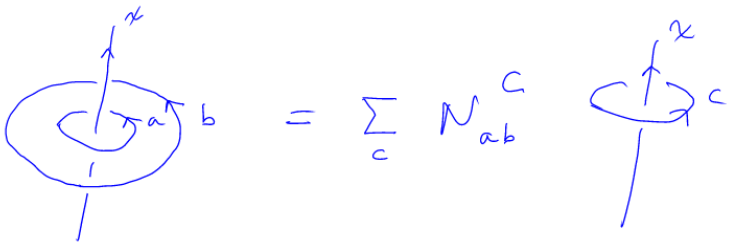


Fig. 12.4

On the other hand, we could also evaluate the left hand side of Fig. 12.4 by applying the identity of Fig. 12.3 twice in a row, and similarly we can evaluate the right hand side of Fig. 12.4 by applying Fig. 12.3 once. 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. 12.2, 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. 8.5!

12.4 Return of Kirby Color

As mentioned in section 12, one can assemble a string called the "Kirby Color" (or Ω string) that is the sum of all strings weighted by the S-matrix.

$$|\Omega\rangle = \sum_{a} S_{0a}|a\rangle = \frac{1}{\mathcal{D}}\sum_{a} d_{a}|a\rangle \tag{12.3}$$

This string has some remarkable properties. Suppose we loop this string around a string x similar to that of Fig. 12.3. 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. 12.5

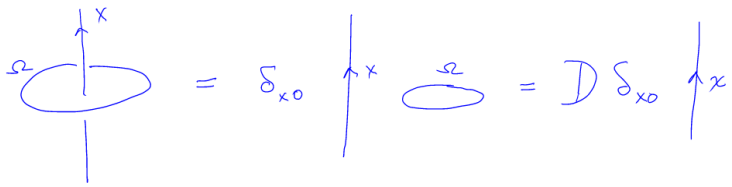


Fig. 12.5 The killing property. A loop of the Kirby color string projects to the vacuum going through it.

Thus, a loop of Kirby color string projects to zero (or vacuum) flux going through it! This is sometimes known as the "killing property", as

a loop of Ω string kills any non-trivial particle that tries to go 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 12 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. See the more detailed discussion in section ***.

12.5 S and θ and that is all?

In building up an anyon theory, we now have compiled a large amount of data. Say there are M particle types, then we have F matrices, which have 6 indices[6], 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 both with one index each. 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 (see section 10.4), 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! This statement is not proven, but there are no counter-examples known.

12.6 Appendix: Quantum Dimensions Satisfy the Fusion Algebra

We would like to show the identity shown in Fig. 12.2. 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. 12.6.

[6]Here we assume no fusion multiplicity greater than 1. If we have such multiplicities, we would add additional indices to the F-matrices.

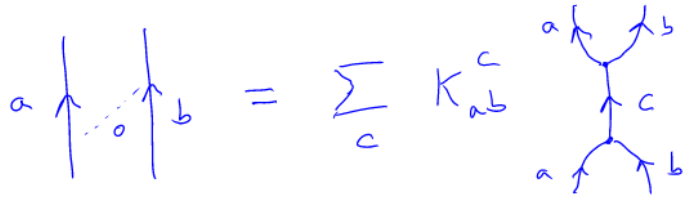


Fig. 12.6 An F-move. If a and b do not fuse to c, then the coefficient κ_{ab}^c must be zero. And if a and b do fuse to c then κ_{ab}^c is not zero. Note that the constant κ_{ab}^c shown here is typically notated as F_{abc}^{ab0}. This is quite similar to $[F_{ab\bar{b}}^a]_{0c}$ except that some lines pointing up have been turned down. This incurs certain normalization factors that one needs to keep track of. See section ***

Further we can use the locality principle (See Fig. 8.7) to give us Fig. 12.7

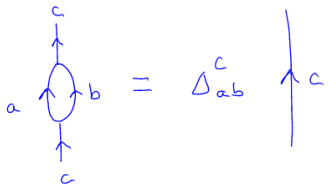


Fig. 12.7 Removal of a bubble gives a factor, which we call $\Delta_{ab}^c \neq 0$.

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. 12.8

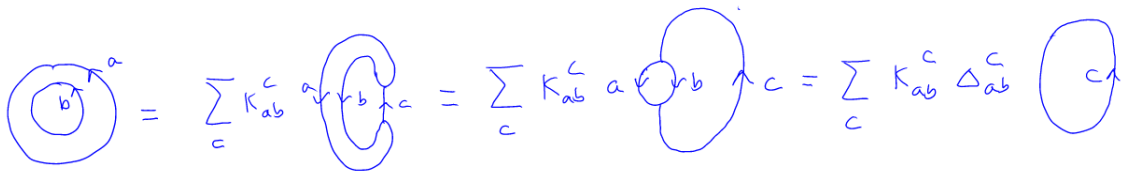


Fig. 12.8 We have applied first the result of Fig. 12.6 then Fig. 12.7. Note that if a and b cannot fuse to c then that term is zero in the sum.

However, we can also apply the same reasoning to split the loops into multiple bubbles as shown in Fig. 12.9.

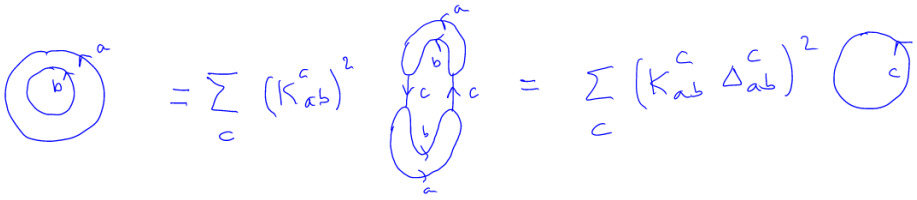


Fig. 12.9 Applying the result of Fig. 12.6 twice then Fig. 12.7 twice.

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 $N_{ab}^c > 1$)[7]. 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. 12.4.

12.7 Appendix: Purely Algebraic Proof of Verlinde Relation

In this section we assume only that we have a set of symmetric fusion matrices $[N_a]_c^b$ which represent the fusion algebra[8]. Nowhere do we need to know anything about the braiding properties of the particle types (indeed, a braiding need not even be defined!). The fusion matrices must be commutative as in Eq. 8.3 so that they are all simultaneously diagonalizable by a unitary matrix which we will call U for now (See Eq. 8.5) which we write as

$$N_a = U \lambda^{(a)} U^\dagger \tag{12.4}$$

where $\lambda^{(a)}$ is a diagonal matrix for each a. Thus the columns of U are eigenvectors of the N matrices which we write as

$$\sum_c [N_a]_b^c U_{cd} = U_{bd} \lambda_d^{(a)}$$

and no sum on d implied. Note, at this point, the columns of U may be multiplied by an arbitrary phase (i.e., a phase redefinition of the eigenvectors).

Since there is a particle type labeled the vacuum 0 (or identity) which fuses trivially with all other particles, we have $[N_a]_0^c = \delta_a^c$ so we have

$$U_{ad} = \sum_c [N_a]_0^c U_{cd} = U_{0d} \lambda_d^{(a)}$$

[7] 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.

[8] The argument of this section is reproduced from Bonderson,Patel, Shtengel, and Simon, to be published.

so that

$$\lambda_d^{(a)} = U_{ad}/U_{0d}$$

substituting back into Eq. 12.4 we get

$$[N_a]_b^c = \sum_x U_{bx} \frac{U_{ax}}{U_{0x}} U_{cx}^*. \tag{12.5}$$

Using Eq. 8.2, we immediately obtain that

$$U_{cx} = U_{\bar{c}x}^*$$

In particular this implies U_{0x} is real. Using Eq. 12.1 (also the result of exercise 8.3)

$$d_a d_b = \sum_c N_{ab}^c d_c$$

we see that the vector $\vec{d}$ must be an eigenvector of N_a for all a, and hence is a column of the matrix U. We choose this to be U_{0x}, and indeed this must be the eigenvector with the largest eigenvalue for each N_a.

Further reading

This is some reading

Exercises

Exercise 12.1 *Kirby Color*

From any anyon theory (i.e., TQFT or modular tensor category) we can construct a type of string (a sum of particle types) called an Ω (sometimes ω) string, or sometimes called a Kirby-color string as given in Eq. 12.3.

(a) Evaluating a knot diagram with the evaluation rules of the TQFT gives

$$Z(S^3 \text{with link})/Z(S^3)$$

So the empty diagram is give value 1.

Consider a simple ring (an "unknot" or unknotted loop of string), blackboard framed (meaning no twists) of Kirby color string. Evaluate this diagram.

(b) A knot (or link) of Kirby-colored string is meant to be equivalent to doing surgury on a the knot thickened into a torus. Considering the result of part (a) above as well as part (a) of the above exercise 7.1 on Surgery. Are these results consistent?

(c) Show that the Ω string made into a loop has the so-called "killing property" shown in Fig. 12.5. In other words, any diagram gives zero unless the particle type going through the Ω loop is the trivial or vacuum particle. Hint:

Use the fact that the quantum dimension is part of the modular S matrix, and various properties of the S matrix to prove this identity.

(d) Evaluate a Hopf Link of Kirby color string (See Fig. 7.21). Does this match the result of part (b) of the exercise 7.1 above?

(e) [Harder] Evaluate the Borromean rings of Kirby color string (See Fig. 7.22). Compare your result to that of part (c) of exercise 7.1 above, and also the discussion in the problem on "Ground State Degeneracy" above.

Hints: Consider the F-move shown in figure 12.10. By closing up the top

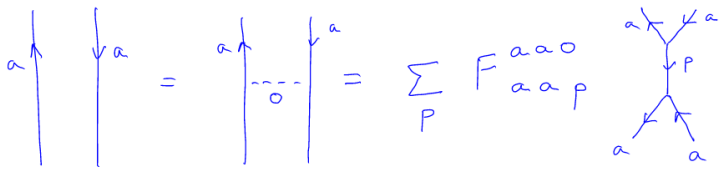


Fig. 12.10 An F-move. The far left diagram can be thought of as having a vacuum particle go between the two strings (middle). Then we can use an F-move to obtain the diagram on the right

and bottom, show that $F_{aa0}^{aa0} = 1/d_a$ with d_a the quantum dimension of the particle a. You will need the locality law (also known as the "no-tadpole" law), which says that diagrams of the type shown in Fig. 12.11 must be zero unless the incoming particle is the vacuum, $p = 0$.

Fig. 12.11 A tadpole diagram must be zero unless $p = 0$, by locality

Exercise 12.2 *Handle Slide* As discussed in section 7.4.2, one can describe a 3-manifold by giving a knot (or link) diagram which should be thickened into a tube and surgered. A handle-slide of a link diagram (which corresponds to sliding a handle of the manifold over another handle, but leaving the manifold topologically unchanged) involves splitting one strand, having it trace the path of a second strand and then reconnecting. An example is shown Fig. 7.19. In TQFT, one uses a string of Kirby color to represent the knot or link to be surgered. In fact, the evaluation of the link in the diagramatic calculus is unchanged by handle-slides. While it takes a bit more diagramatic calculus rules to derive the handle-slide invariance in general, a simple case of the handle-slide is fairly easy to derive. Consider instead a handle-slide over an untwisted loop as shown in the figure 12.12. Use the killing property. You will have to think about fusion, but you should not need to do any detailed calculations with F matrices.

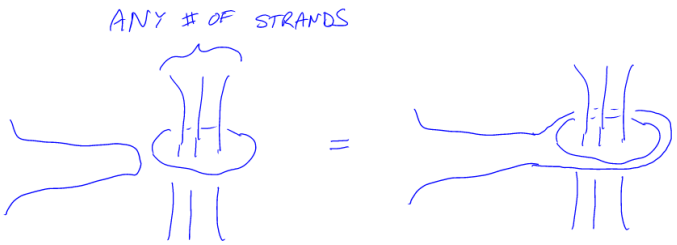


Fig. 12.12 A handle-slide over an untwisted loop. In this figure all strings are meant to be Ω strings or Kirby color (a weighted sum of all particle types).

Exercise 12.3 *Fusion and Ground State Degeneracy* In exercise 8.2 above, we have calculated the ground state degeneracy for a TQFT on an arbitrary oriented 2d manifold Σ. Using the Verlinde relation, show that the ground state degeneracy can be written as

$$\dim = Z(\Sigma \times S^1) = \sum_x [S_{0x}]^\chi$$

where $\chi = 2-2g$ is the Euler characteristic where g is the *genus* of the manifold (the number of handles).

Quantum Error Correction and The Toric Code

13

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.

13.1 Classical Versus Quantum Information

13.1.1 Memories[1]

[1] All alone in the moonlight!

Classical Memory

The unit of classical information is a bit — a 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 two-dimensional complex Hilbert space spanned by vectors which we usually call $|0\rangle$ and $|1\rangle$. A qubit can be in any state

[2] Sometimes q-bit, but never cubit.

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$$

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

$$|\psi\rangle = \alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle \tag{13.1}$$

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!

13.1.2 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.

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). Here we use three "physical" bits to store one "logical" bit of information.

logical bit	physical bits
0	000
1	111

Table 13.1 Three bit repetition code. Stores a single logical bit of information using three physical bits.

Our memory should either be in 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 physical bits, such as 5 physical bits or 7 physical bits for a single logical bit.

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 (we allow some arbitrary phase χ in the copying process).

$$\begin{aligned} U(|0\rangle \otimes |\phi_2\rangle) &= e^{i\chi}|0\rangle \otimes |0\rangle \\ U(|1\rangle \otimes |\phi_2\rangle) &= e^{i\chi}|1\rangle \otimes |1\rangle \end{aligned}$$

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 a clone of the bit should have given the outcome

$$e^{i\chi}[\alpha|0\rangle + \beta|1\rangle] \otimes [\alpha|0\rangle + \beta|1\rangle]$$

which is not generally the same result. Thus no cloning device is consistent with the linearity inherent in quantum mechanical evolution.

13.2 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).

[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. To evaluate the quality of a code one must make reasonable assumptions about how likely a physical qubit is to fail and compare this to how quickly one can test for errors and correct them. NEED CITATION HERE?

13.2.1 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 "toric"). The total number of spins is $N = 2N_xN_y$ and correspondingly the dimension of the Hilbert space is 2^N.

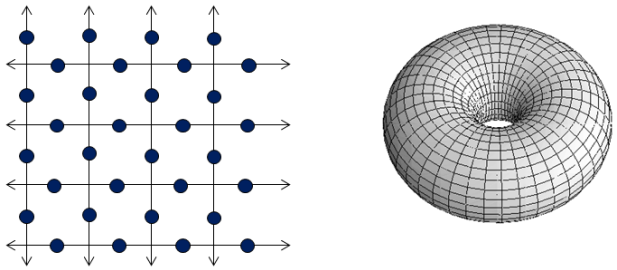


Fig. 13.1 The Hilbert space of the toric code — an N_x by N_y square lattice with spins (dots) on each edge wrapped up to make it periodic in both directions — i.e., a torus. Hence the name. There are 32 spins in this picture so the Hilbert space has dimension 2^{32}.

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. 13.2.

[4]Caution: In the literature about half of the world uses the up-down or σ_z eigenstates as a basis, and half of the world uses the σ_x eigenstates as a basis.

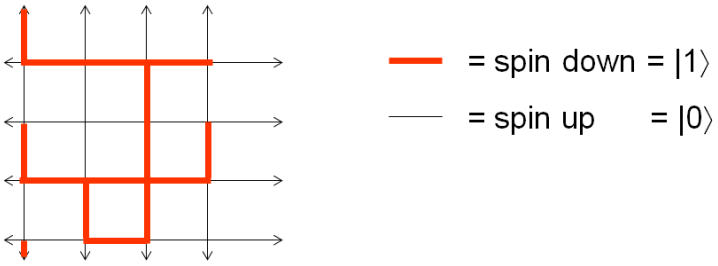


Fig. 13.2 A particular basis state of the Hilbert space, working in the up-dpwn basis (z-eigenstates). Here we denote down spins by thick (red) lines. And up spins are denoted by not coloring in the edges.

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. 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.

13.2.2 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. 13.3. Note that there are a total of $N_x N_y$ vertex operators.

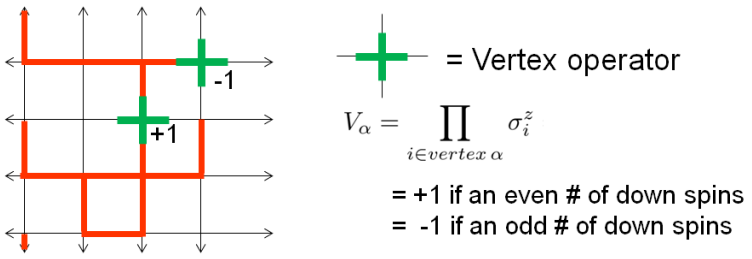


Fig. 13.3 The vertex operator returns $+1$ if there are an even number of incident down spins at that vertex and returns -1 if there are an odd number.

Note that it is possible (and useful) to define a corresponding projection operator

$$\tilde{V}_\alpha = \frac{1}{2}(1 - V_\alpha) \tag{13.2}$$

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. 13.4. There are a total of $N_x N_y$ plaquette operators.

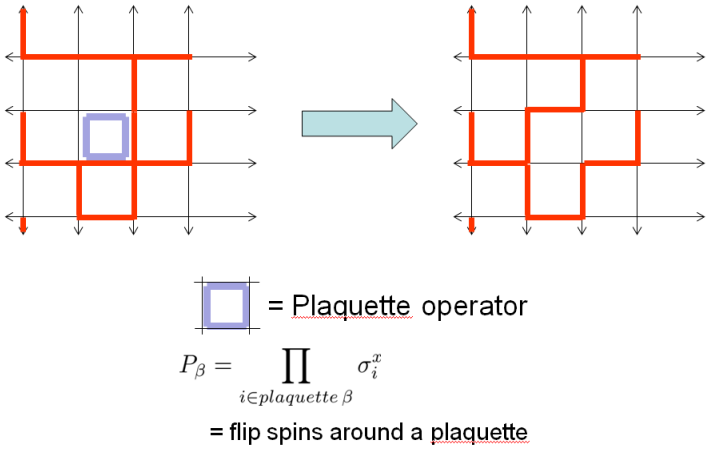


Fig. 13.4 The plaquette operator flips the state of the spin on the four edges of a plaquette.

As with the vertex operator, $P_\beta^2 = 1$ meaning P_β has eigenvalues $+1$ and -1. We can similarly define a projector

$$\tilde{P}_\beta = \frac{1}{2}(1 - P_\beta) \tag{13.3}$$

which satisfies $P_\beta^2 = P_\beta$.

It is a bit more difficult to describe what the 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. 13.5. The orthogonal superposition (adding the two states with a - sign) will give the other eigenstate.

Fig. 13.5 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 $2N_x N_y$ 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 $(N_x N_y - 1)$ operators V_α, but then the value of the one remaining V_α is then fixed by the values chosen for the other $(N_x N_y - 1)$ of them. Similarly with the P_β's. So specifying the eigenvalues of these commuting operators specifies only $2(N_x N_y - 1)$ degrees of freedom, and since we started with $2N_x N_y$ 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.

13.2.3 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. 13.6

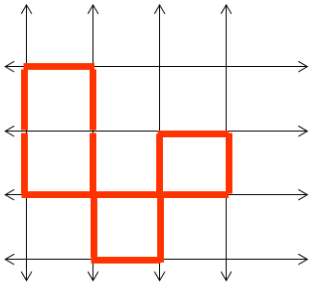


Fig. 13.6 A loop configuration consistent with the constraint that $V_\alpha = 1$ on every vertex. There must be an even number of red lines incident on every vertex.

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. 13.5 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

$$|\psi\rangle = \sum_{\substack{\text{all loop configs that can be obtained by flipping plaquettes from a reference loop config}}} |\text{loop config}\rangle \qquad (13.4)$$

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. 13.7. 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.

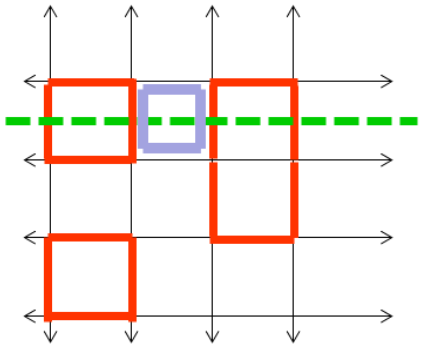


Fig. 13.7 Making a cut around one of the handles of torus, one can see that flipping a plaqutte, such as the blue one, does not change the parity of the number of red bonds cutting the green line. Further, it does not matter where (at which y-coordinate) the green cut is made, the number of red bonds it cuts is always even.

Thus there are four independent wavefunctions of the form of Eq. 13.4, 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. 13.8

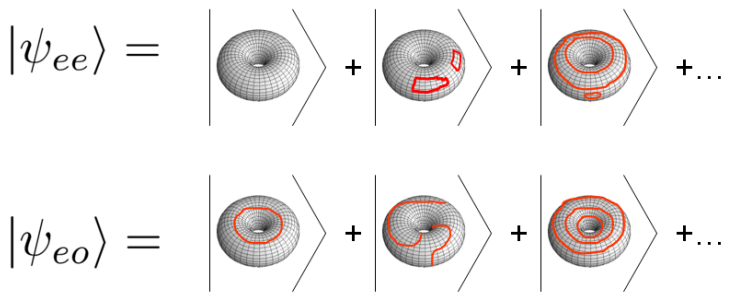


Fig. 13.8 Graphical depiction of $|\psi_{ee}\rangle$ which has an even number of strings running around each handle, and $|\psi_{eo}\rangle$ which is even around the first handle odd around the second.

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

$$|\psi\rangle = A_{ee}|\psi_{ee}\rangle + A_{eo}|\psi_{eo}\rangle + A_{oe}|\psi_{oe}\rangle + A_{oo}|\psi_{oo}\rangle \tag{13.5}$$

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. 13.1). We will refer to wavefunctions of the form of Eq. 13.5 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.

13.3 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.

13.3.1 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. 13.9.

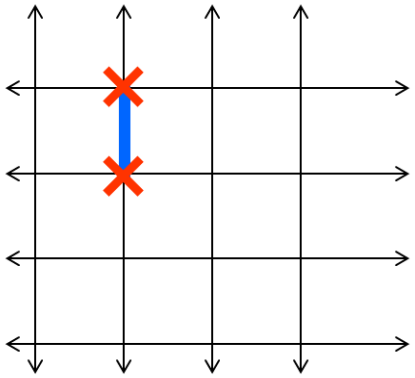


Fig. 13.9 A σ^x operator applied to the bond creates two vertices in the $V_\alpha = -1$ eigenstate.

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. 13.10, 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.

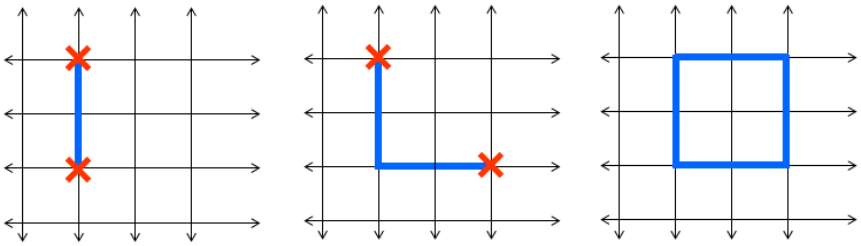


Fig. 13.10 Left: When two σ^x errors are made on bonds that share a vertex, the shared vertex is hit with σ_i^x twice, and thus becomes $V = +1$ again. Only the two vertices at the end of the "string" are in the $V = -1$ state. **Middle:** A longer string of errors. Note that we can only measure the 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 13.10 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. 13.10, 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. 13.10. 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.

13.3.2 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. 13.11. Recall that this eigenstate of the plaquette operator is a superposition of the flipped and unflipped plquettes similar to that shown in Fig. 13.5 but with a minus sign between the two terms.

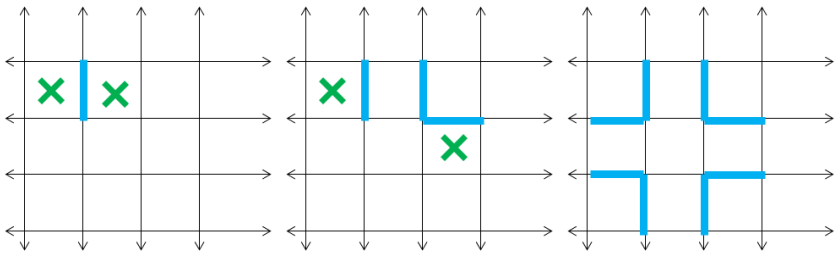


Fig. 13.11 Left: When a σ^z error is applied to a bond, the plaquettes on either side end up in the $P = -1$ state **Middle:** A string of several σ^z errors. **Right** A closed loop of σ^z errors. This is equal to the product of all of the enclosed V_α operators. In the code space, this is equal to $+1$.

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. 13.11. 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. 13.11, 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. 13.12, this operator then counts the parity of the number of red bonds going around the dual handle, as shown in the figure.

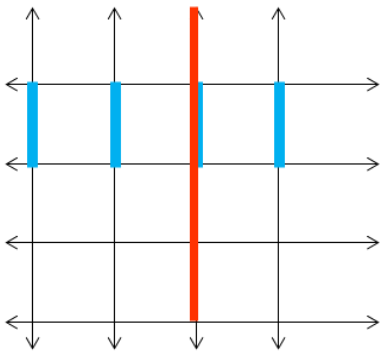


Fig. 13.12 If a string of σ^z goes around a handle, it measures the parity of the number of red strings going around the dual handle.

Thus, applying the string of σ^z operators around the handle makes the transformation

$$A_{ee}|\psi_{ee}\rangle + A_{eo}|\psi_{eo}\rangle + A_{oe}|\psi_{oe}\rangle + A_{oo}|\psi_{oo}\rangle \longrightarrow$$
$$A_{ee}|\psi_{ee}\rangle + A_{eo}|\psi_{eo}\rangle - A_{oe}|\psi_{oe}\rangle - A_{oo}|\psi_{oo}\rangle$$

13.3.3 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!

13.3.4 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.

13.4 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

$$H = -\sum_{vertices\,\alpha} V_\alpha - \sum_{plaquettes\,\beta} P_\beta \tag{13.6}$$

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. 13.2 and 13.3. Writing

$$\tilde{H} = \sum_{vertices\,\alpha} \tilde{V}_\alpha + \sum_{plaquettes\,\beta} \tilde{P}_\beta \tag{13.7}$$

which differs from Eq. 13.6 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.

13.4.1 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

13.4.2 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.

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

[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. 2.6 or 10.1. Considering Fig. 13.13

shown in the right panel of Fig. 13.10 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. 13.11 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

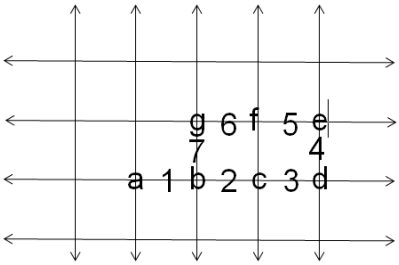


Fig. 13.13 Vertices are labeled with letters and bonds are labeled with numbers.

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. 2.6 or 10.1. 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.

phase under a twist as shown in Fig. 13.14

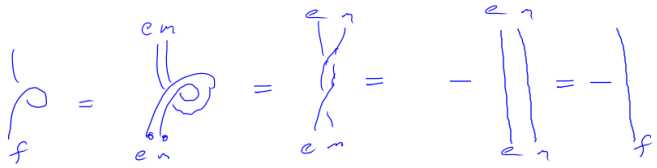


Fig. 13.14 The $f = e \times m$ particle is a fermion, since e braiding around m gives a -1 sign.

Note that taking f all the way around f will result in a net + sign.

13.4.3 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.7.13. 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 $\mathcal{D} = 2$.

13.4.4 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 magneti

It is easy to see that this charge and flux will correctly give the +1 and -1 phases accumulated from braiding particles.

13.5 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 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

[6]This is a Brillouin-Wigner perturbation theory, where successive terms are rigorously λ/Δ smaller.

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 10.4 — it is not possible that the braiding phases in a theory change an arbitrarily small amount.

13.6 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.

Kitaev's Generalized Toric Code: The Quantum Double of a Group — Lattice Gauge Theory

14

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. 14.1

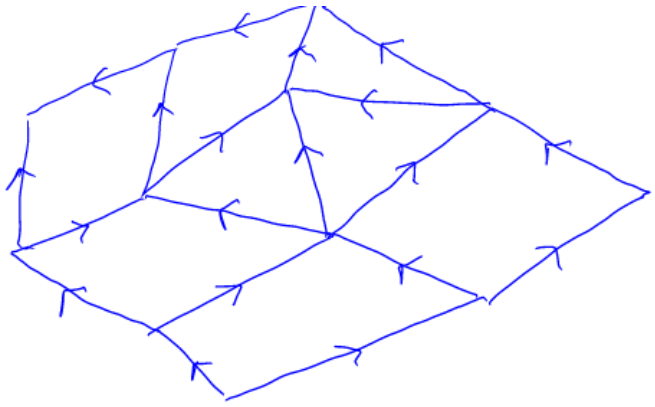


Fig. 14.1 Part of a directed graph.

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. 14.2.

[1] I present this model on the "dual" graph compared to Kitaev's presentation.

Fig. 14.2 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. 14.3. This is the string-net vertex fusion rule.

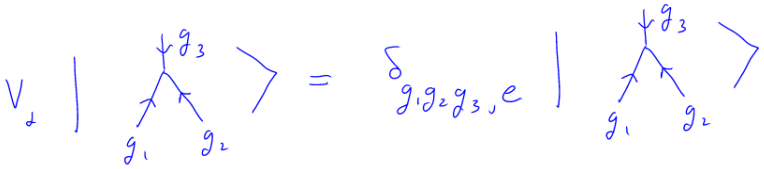


Fig. 14.3 Definition of V_α when all arrows are directed into the vertex (if a vertex is directed out, one can invert the arrow and invert the group element). The vertex operator gives zero unless the product of group elements around the plaquette gives the identity element e

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. 14.4.

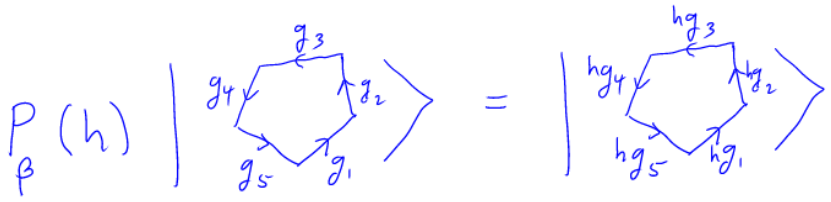


Fig. 14.4 The plaquette operator $P_\beta(h)$ premultiplies all of the clockwise oriented bonds by the element h.

The total plaquette operator that (the one that will enter the Hamil-

tonian) 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.

$$\begin{aligned} P_\beta(1) &= \text{identity operator} \\ P_\beta(-1) &= \text{multiply all edges by -1. (i.e. flip all edges)} \end{aligned}$$

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

[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 $ug = gu$.

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.

More Generalizing The Toric Code: Loop Gases and String Nets

15

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.

15.1 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. 15.1

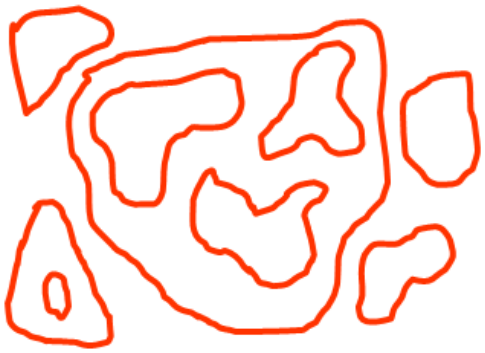


Fig. 15.1 A loop gas in 2d. We can think of this as particle world-lines in 1+1 d.

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

$$|\psi\rangle = \sum_{\substack{\text{all loop configs that can be} \\ \text{obtained from a reference} \\ \text{loop config}}} |\text{loop config}\rangle \tag{15.1}$$

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. 15.2. We have always allowed smooth deformations in our diagrammmatic algebras.

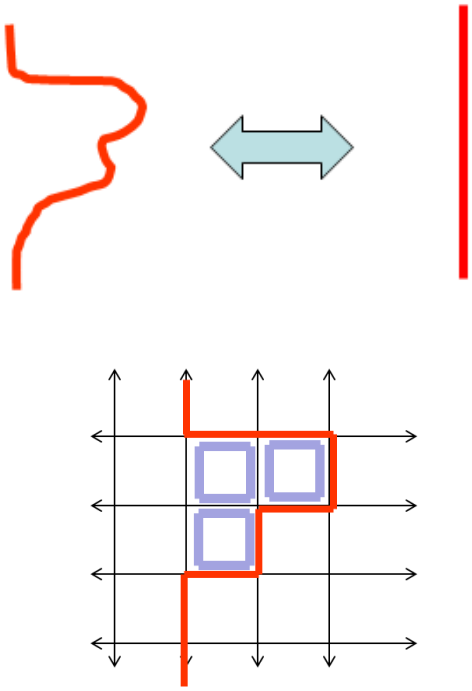


Fig. 15.2 Isotopy (Top) Off the lattice this is just deformation of a line. (Bottom) on the lattice, this is implemented by flipping over the blue plaquettes.

Move 2: "Adding or removing a loop". As shown in Fig. 15.3

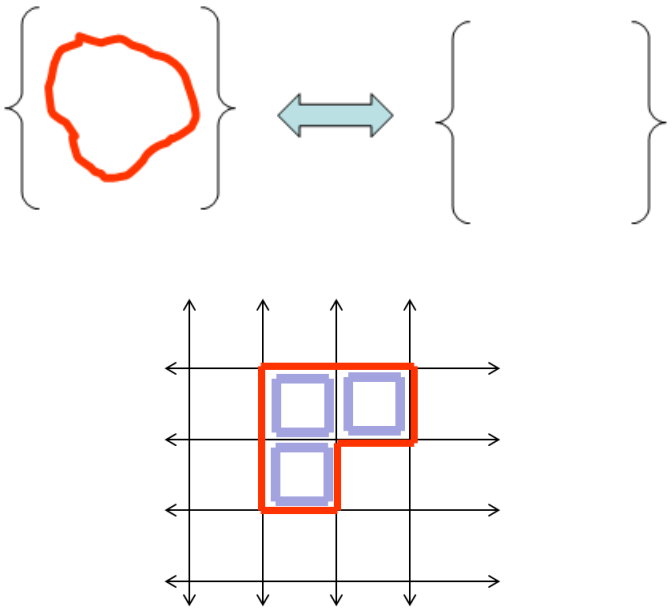


Fig. 15.3 Adding or Removing a loop (Top) Off the lattice (Bottom) On the lattice we flip the shown plaquettes.

Move 3: "Surgery" or reconnection of loops. As shown in Fig. 15.4

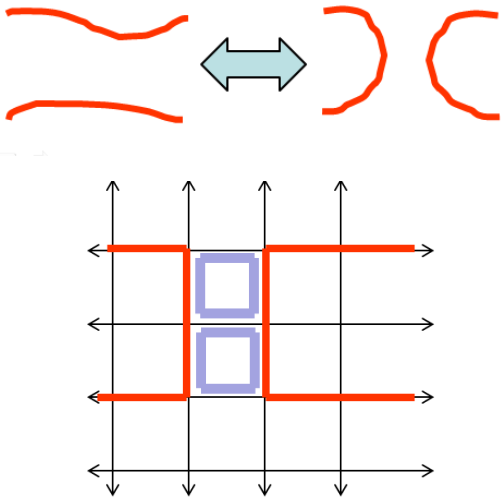


Fig. 15.4 Loop Surgery (Top) Off lattice surgery (Bottom) On lattice, flip the shown plaquettes

We can summarize these rules with simple skein-like relations as shown in Fig. 15.5

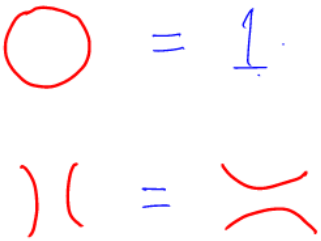


Fig. 15.5 "Skein" relations for the toric-code loop gas. The unity on the right of the top line means that the amplitude in the superposition that forms the wavefunciton is unchanged (multiplied by unity) under removal or addition of a loop.

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.

15.1.1 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. 15.6.

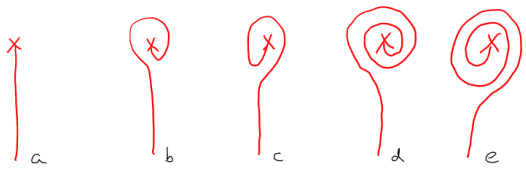


Fig. 15.6 Ends of strings can be wrapped either way, and multiple times. a and b are different, c is equivalent to b by surgery. Similarly d and e are both the equivalent to a.

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

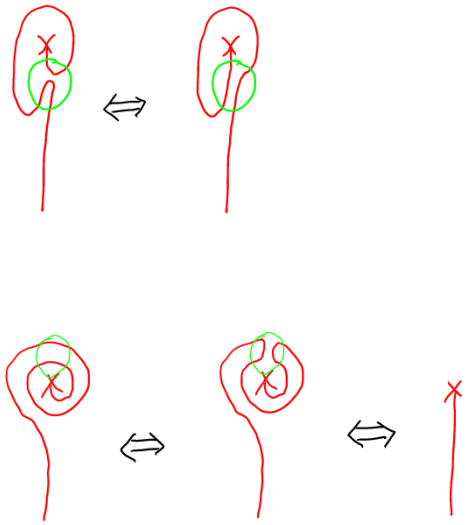


Fig. 15.7 Loop equivalences. Surgery is done inside the light green circles. The final equality on the lower right is just pulling the string tight.

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. 15.8

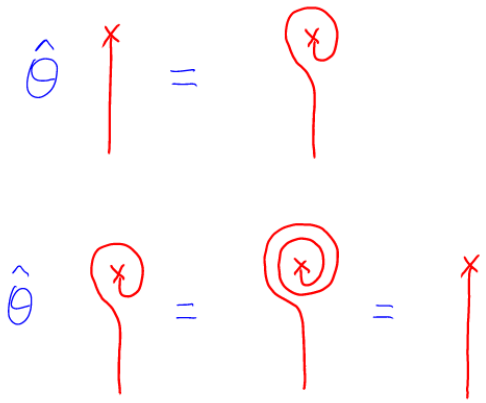


Fig. 15.8 Rotation

From these relations we can determine that the eigenvalues of the rotation operator are $+1$, correpsonding to the e particle and -1 corre-

sponing to the f particle, as shown in Fig. 15.9.

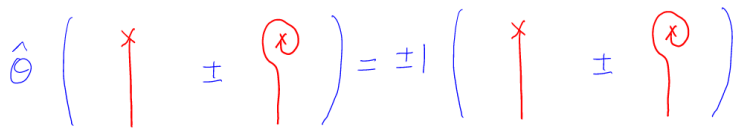


Fig. 15.9 The eigenvectors of the rotation operator $\hat{\theta}$

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. 15.10.

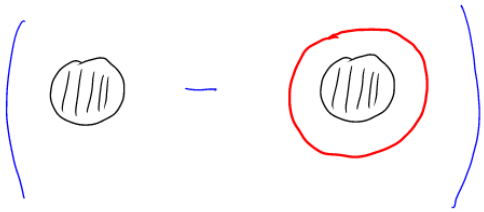


Fig. 15.10 The black disk is some region of our model. Forming a superposition of this region, and this region with a loop around it, with a minus sign between the two pieces, must be orthogonal to the ground state — it puts a magnetic excitation m in the region.

15.2 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.

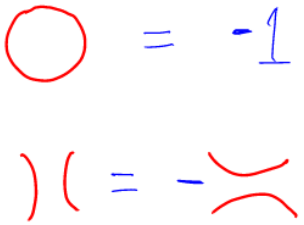


Fig. 15.11 "Skein" relations for the double-semion loop gas. Each loop removal/addition and each surgery incurs a minus sign. Note that these are the same as the Kauffman rules when we considered semions.

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

$$|\psi\rangle = \sum_{\substack{\text{all loop configs that can be}\\ \text{obtained from a reference}\\ \text{loop config}}} (-1)^{\text{Number of Loops}} |\text{loop config}\rangle \tag{15.2}$$

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.

15.2.1 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.15.12). 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. 15.12. The advantage of this structure is that loops cannot intersect as they can (at the 4-fold corner) on the the square lattice.

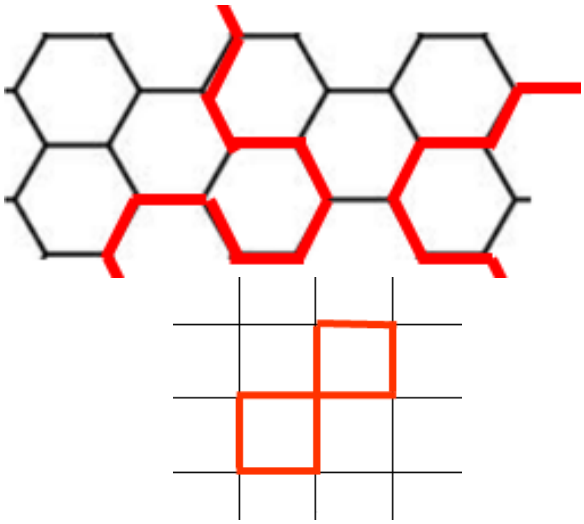


Fig. 15.12 Left: Toric code on a honeycomb, loops are nonintersecting. Right: On the square lattice loops can intersect at corners and one cannot tell if this picture represents one loop or two.

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. 15.13. Depending on the initial state, when the plaquette is flipped, one may or may not obtain a minus sign.

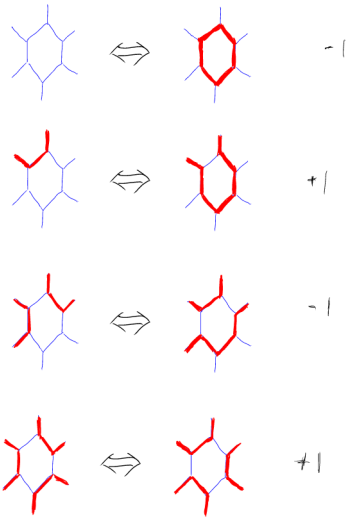


Fig. 15.13 Some plaquette flips for the double semion model on the hexagon. The top line obviously adds a loop, so should get a minus sign. The second line just stretches a loop over a plaquette, so does not get a minus sign. The third line is a surgery so gets a minus sign. The fourth line is a double 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.

15.2.2 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. 15.14

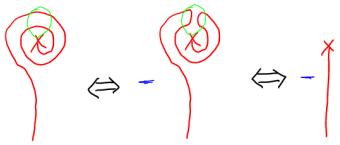


Fig. 15.14 Surgery incurs a minus sign. Compare to fig. 15.7

Resulting in the effect of rotation being Fig. 15.15

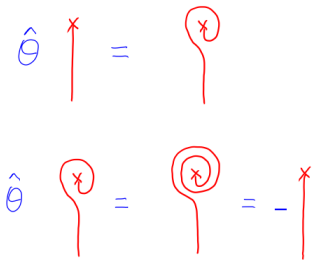


Fig. 15.15 Surgery incurs a minus sign. Compare to fig. 15.7

Again we can use these to give us the eigenstates of the rotation operator as shown in Fig. 15.16

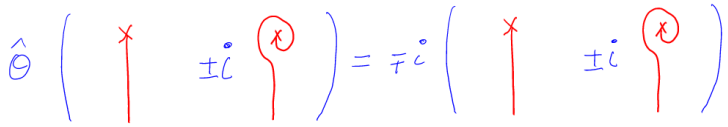


Fig. 15.16 Eigenstates of the rotation operator for the doubles semion model.

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. 15.10, 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. 15.10 (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.

15.3 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. 15.17. 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!).

[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 12.4, and gives $\mathcal{D}$ on the vacuum.

Fig. 15.17 A general string net, that allows branching, here with two colors.

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. 15.18.

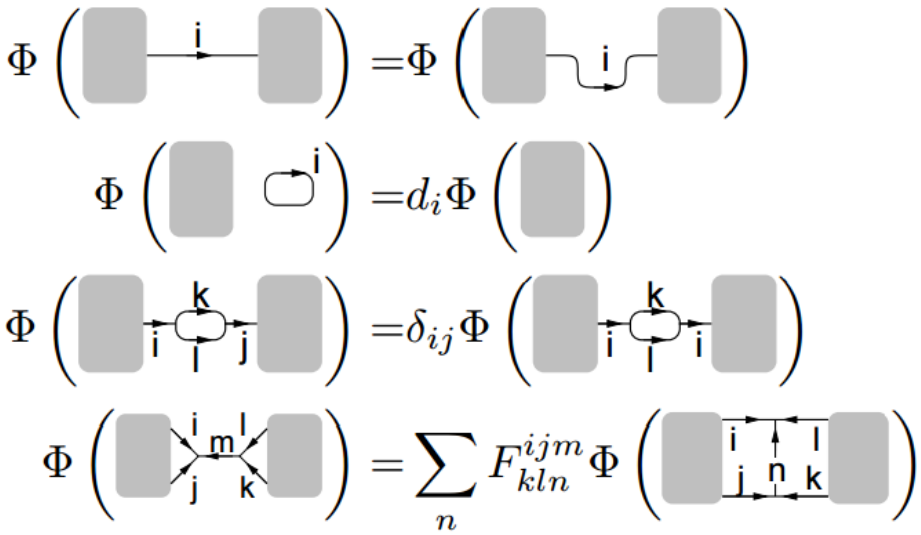


Fig. 15.18 Rules for a string net. The grey regions are meant to be the same on both the left and the right of the diagram. Figure stolen from Levin and Wen.

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. 15.5 and Fig. 15.11.

It is important to note that the F-matrix used to define define the string net (last line of Fig. 15.18) 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.

15.4 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. 15.19.

Fig. 15.19 A branching string net for the doubled Fibonacci model.

Starting with Eq. 8.6, we consider the following F-moves as shown in Fig. 15.20

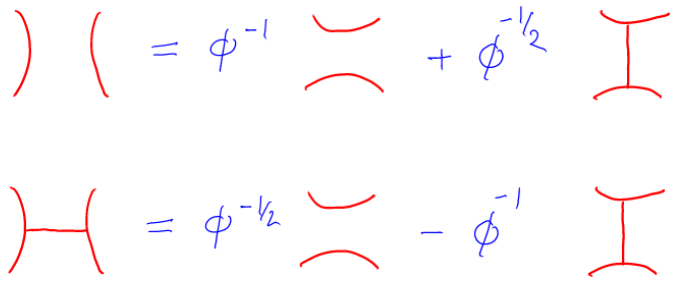


Fig. 15.20 Rules for building the doubled fibonacci model.

Where here $\phi = (1+\sqrt{5})/2$ and (the values of these coefficients come from the values of the F-matrix in Eq. 8.6.

We also expect to have rules of the form of Fig. 15.21

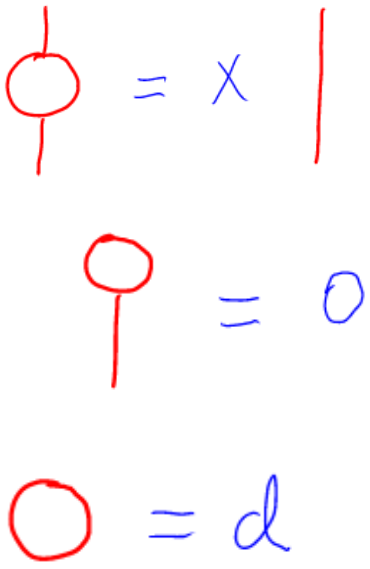


Fig. 15.21 Rules for building the doubled fibonacci model.

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. 15.20 to give Fig. 15.22.

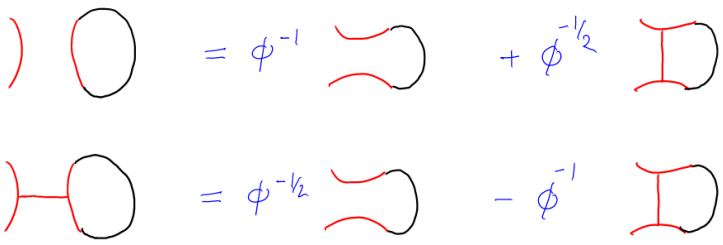


Fig. 15.22 Starting with Fig. 15.20 and closing strings to the right hand. The black strings should be imagined to be red — they are drawn black so one can see what is added compared to Fig.!15.20

Using the laws above we these equations are translated to

$$\begin{aligned} d &= \phi^{-1} + \phi^{-1/2} X \\ 0 &= \phi^{-1/2} - \phi^{-1} X \end{aligned}$$

which we solve to obtain

$$\begin{aligned} X &= \phi^{1/2} \\ d &= \phi^{-1} + 1 = \phi \end{aligned}$$

The fact that $d = \phi$ is not surprising being that this is the expected quantum dimension for a Fibonacci particle.

With the values we obtain for X and d, we now have a full set of rules in Fig. 15.20 and 15.21. 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

[2] In fact we can prove that the tadpole rule must be zero. This is a homework problem!

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 15.21) 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. 15.20.

15.4.1 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. 15.23

Fig. 15.23 Possible string endings in the doubled fibonacci string net model.

Just as an example, consider the ending shown on the left of Fig. 15.24. By using an F-move, it is reduced to a combination of the three presented above.

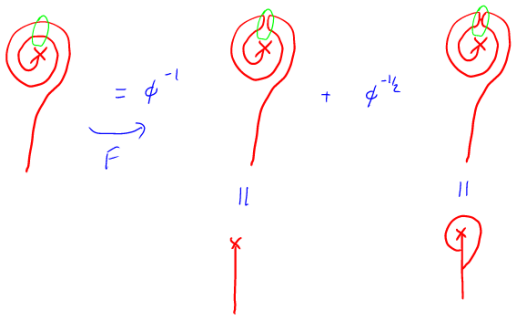


Fig. 15.24 An example of reducing a more complicated string ending into one fo the three endings shown in Fig. 15.23.

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. 15.25 and then using F-matrices to reduce the result back to linear combinations of the same three possible endings.

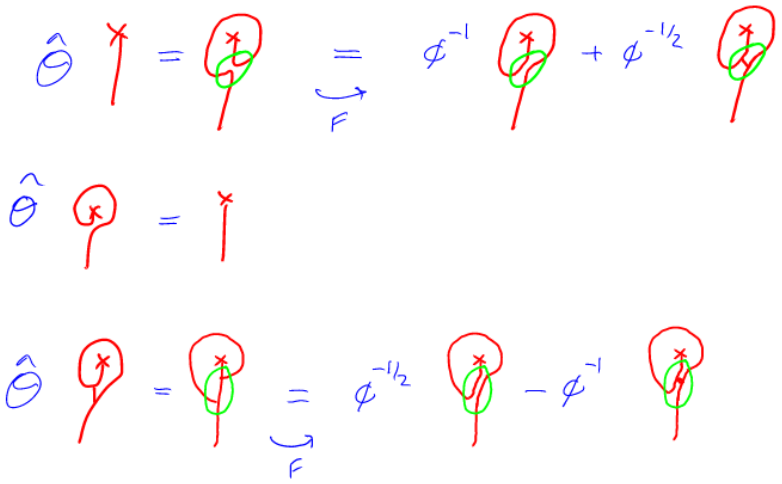


Fig. 15.25 The rotation operator $\hat{\Theta}$ applied to the possible string endings. Then using F matrices we reduce the results to linear combinations of the same endings.

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 10.3.). 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!).

15.4.2 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. 15.26

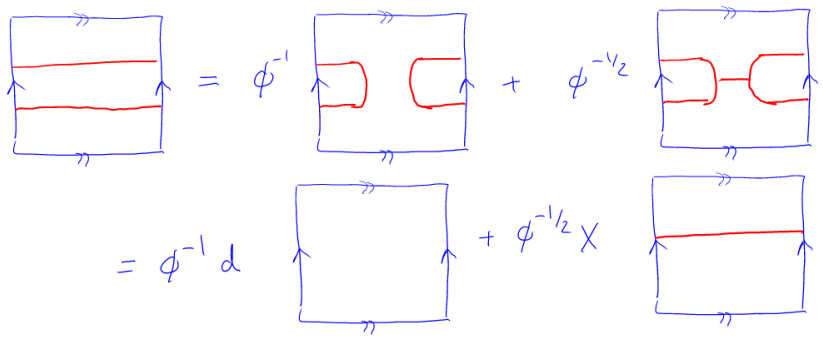


Fig. 15.26 Reducing two loops around a handle to a linear combination of one loops and zero loops.

15.5 Add details of Levin Wen Model on the lattice?

15.6 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. 15.27

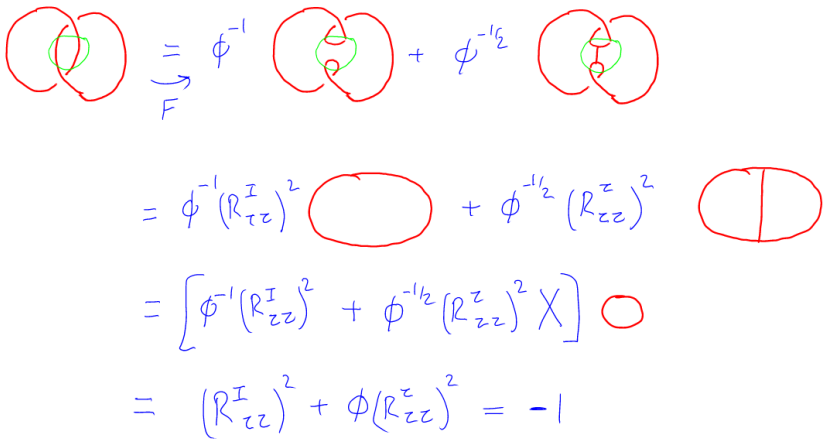


Fig. 15.27 Calculating the nontrivial element of the Fibonacci anyon S-matrix.

Introduction to Quantum Hall — The Integer Effect

16

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]

16.1 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. 16.1). 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.

[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.

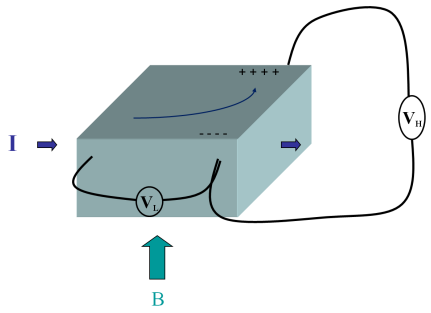


Fig. 16.1 Hall voltage V_H perpendicular to both magnetic field and current, and proportional to both. Also one measures a longitudinal voltage in the same direction as the current, roughly independent of magnetic field.

16.2 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. 16.2.

[3]Metal Oxide Semiconductor Field Effect Transistors

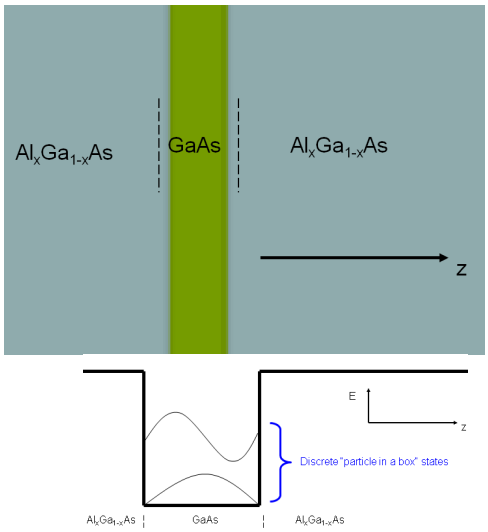


Fig. 16.2 **Top** A quantum well structure is a quasi-two-dimensional layer of one semiconductor sandwiched between two other semiconductors. **Bottom** The potential felt by an electron is like a particle in a box. If the energy is low enough, the electron is stuck in the lowest particle-in-box wavefunction $\varphi_0(z)$ 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. 16.2. 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).

16.3 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

[4]Meaning very clean

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. 16.3.

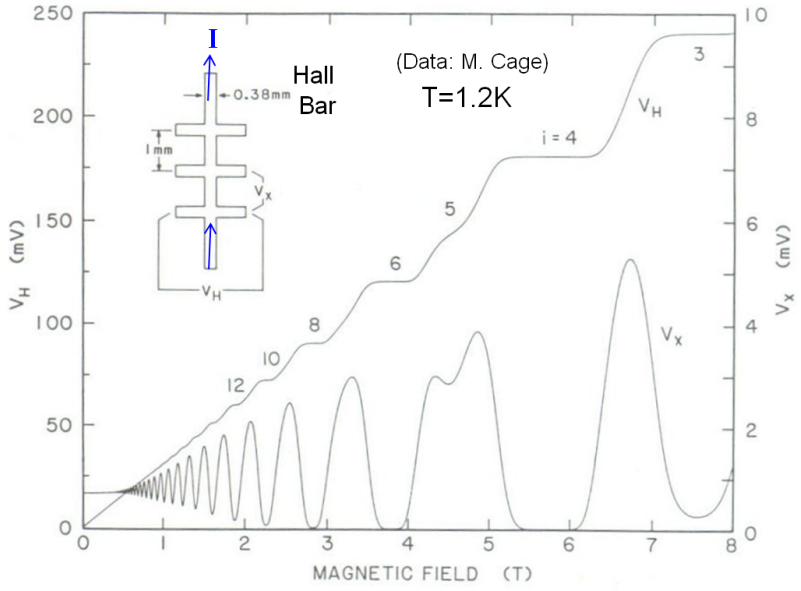


Fig. 16.3 An example of an Integer Quantum Hall experiment. The plateaus in V_H are such that $V_H = (1/i)(h/e^2)I$ with i the integer displayed over the plateau — where h is Planck's constant and e is the electron charge. At the same magnetic field where a plateau occurs in V_H the longitudinal voltage 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

$$R_H = \frac{1}{i}\frac{h}{e^2} = 1/G_H \tag{16.1}$$

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

$$\nu = \frac{n\phi_0}{B}$$

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. 16.1 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.

16.4 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

$$\mathbf{F} = -e\,(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

[5]These are 2 by 2 matrices because they relate the vector electric field $\mathbf{E}$ to the vector current $\mathbf{j}$

[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.

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

$$\begin{aligned} R_L &= 0 \\ R_H &= \frac{B}{ne} \end{aligned}$$

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.

16.5 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 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

$$\left(\frac{p_x^2}{2m} + \frac{1}{2}m\omega_c^2(k_y\ell^2 + x)^2\right)\phi_{k_y}(x) = E\phi_{k_y}(x) \tag{16.2}$$

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

$$E_p = \hbar\omega_c\left(p + \frac{1}{2}\right) \tag{16.3}$$

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. 16.4.

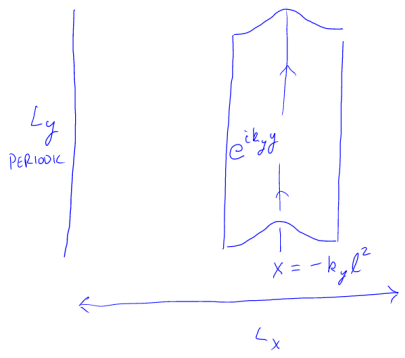


Fig. 16.4 The shape of the wavefunction of an electron in a magnetic field using Landau gauge. The form of the wavefunction will be harmonic oscillator on the x direction and plane-wave in the y-direction

Fixing the energy by fixing p in Eq. 16.3, 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

$$\text{Number of states in a Landau level} = \frac{L_x L_y}{2\pi\ell^2} = \frac{\text{Area}}{\phi_0} B$$

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. 16.5

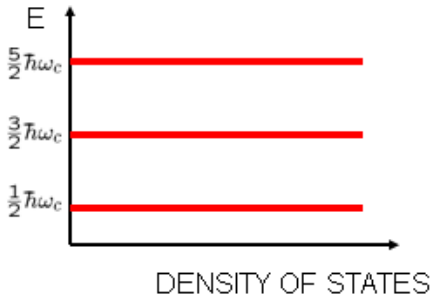


Fig. 16.5 The density of states for spin-polarized (or spinless) electrons in a magnetic field. At energies equal to half-odd integer multiples of the cyclotron frequency, there is a spike of degenerate states, with degeneracy $\frac{\text{Area}}{\phi_0} B$.

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. 16.6.

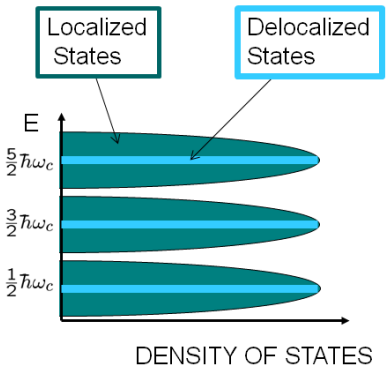


Fig. 16.6 The density of states for spin-polarized (or spinless) electrons in a magnetic field with disoder. The Landau bands are spread out, with localized eigenstates in the tails and extended eigenstates near the middle.

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. 16.6.

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.

16.6 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.

[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.

16.6.1 Byers and Yang Theorem

Consider any system (made of electrons and protons and neutrons) with a hole cut in it, as in Fig. 16.7.

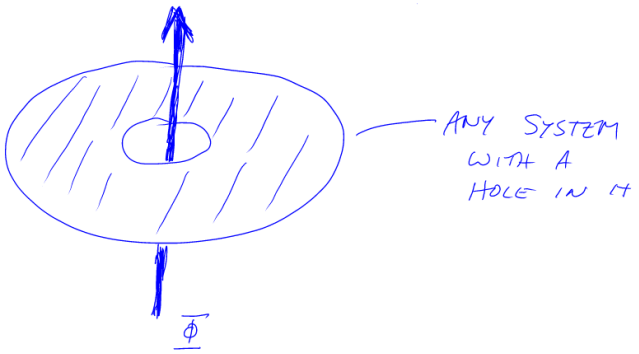


Fig. 16.7 The Byers-Yang theorem states that threading any integer number of flux quanta through a hole in a system leaves the eigenspectrum unchanged.

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).

To prove this theorem we use gauge invariance. One is always free to make a gauge transformation

$$\begin{aligned} \mathbf{A}'(\mathbf{r}) &= \mathbf{A}(\mathbf{r}) + (\hbar/e)\nabla\chi(\mathbf{r}) \\ \Psi'(\mathbf{r}_1, \ldots \mathbf{r}_N) &= \left[\prod_{j=1}^{N} e^{i\chi(\mathbf{r}_j)}\right] \Psi(\mathbf{r}_1, \ldots \mathbf{r}_N) \end{aligned}$$

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,

[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.

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.

16.6.2 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. 16.8.

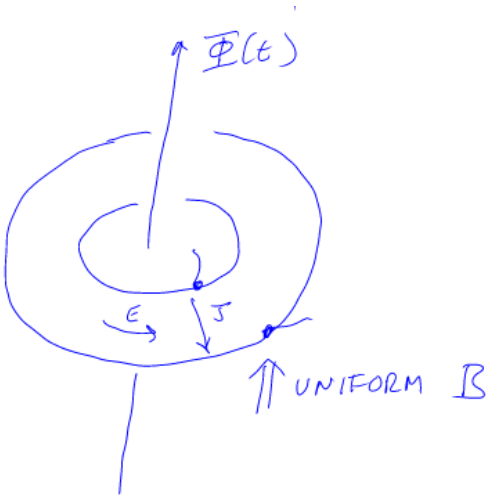


Fig. 16.8 Insertion of Flux $\Phi(t)$ through the center of an annulus of two-dimensional electrons in a uniform magnetic field. Adiabatically increasing the flux creates an electric field in the annular direction which then, by the Hall conductivity, creates current in the radial direction.

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.

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

[11]For studying current flow in magnetic fields, the annulus is knowni as "Corbino" geometry, after O. M. Corbino, who studied this in 1911.

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).

16.6.3 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. 16.9.

[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.

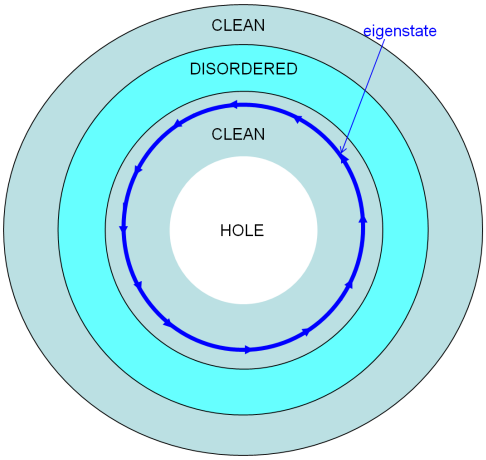


Fig. 16.9 The Halperin geometry. The same as the Laughlin annulus geometry, except here we add disorder only in part of the annulus. We have also shown (dark blue) a single particle eigenstate in the clean region, which forms a circle (with a small gaussian cross-section).

Here, the disorder is confined to only part of the annulus, the innermost 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

[13] Just find the maximum of $|\psi_m|^2$.

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).

Introduction to Fractional Quantum Hall Effect

17

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

[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!

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 (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

$$\overset{\text{insert } \phi_0}{\longrightarrow} |GS_1\rangle \overset{\text{insert } \phi_0}{\longrightarrow} |GS_2\rangle \overset{\text{insert } \phi_0}{\longrightarrow} |GS_3\rangle \overset{\text{insert } \phi_0}{\longrightarrow} |GS_1\rangle \overset{\text{insert } \phi_0}{\longrightarrow}$$

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

[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.

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.

17.0.1 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 16.5, 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. 16.5, 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 (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!).

[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.

[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!

17.1 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 16.6.3) take the form[7]

$$\varphi_m(z) = C_m z^m e^{-|z|^2/(4\ell^2)} \tag{17.1}$$

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. 17.1. 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.

[7] We will ignore the spin degree of freedom as before.

17.1.1 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!

[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.

17.2 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!)

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.

[9]Note that this wavefunction is not normalized in any sense. The issue of normalization becomes important later.

[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

17.2.1 Exact statements about Laughlin Wavefunction

It turns out that the Laughlin wavefunciton is actually the exact ground state of a special inter-particle 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. 17.1.

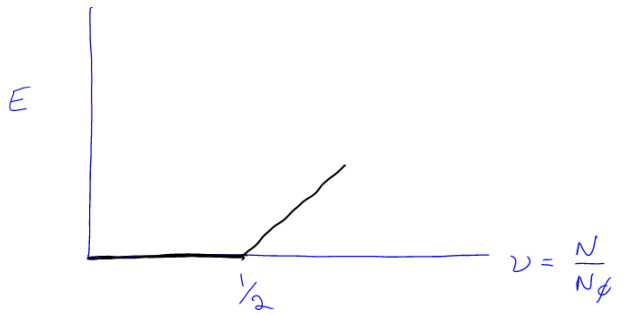


Fig. 17.1 Schematic of the ground state energy as a function of filling fraction for bosons with delta function interaction.

The key point is that the ground state energy has a cusp, which means there is a jump in the chemical potential

$$\mu = \frac{\partial E}{\partial N}$$

[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.

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

$$\Psi(\mathbf{dr}_1 \ldots \mathbf{dr}_N) = \phi(z_1 \ldots z_N) \prod_{i=1}^{N} e^{-|z_i|^2/(4\ell^2)} \tag{17.2}$$

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

$$E = \sum_{i<j} \int \mathbf{dr}_1 \ldots \mathbf{dr}_N \; |\partial_{z_i} \phi|^2 \;\; \delta(\mathbf{r}_i - \mathbf{r}_j) \prod_{i=1}^{N} e^{-|z_i|^2/(2\ell^2)}$$

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].

[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.

[15] Note that by antisymmetry the wavefunction must go as an odd number of powers as two particle positions approach each other.

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.

17.2.2 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].

17.3 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.*

17.3.1 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 precisly one empty orbtial open at positon $\mathbf{0}$, thus leaving a positive charge

$$e^* = \nu e$$

[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.

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 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).

17.3.2 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. 17.2 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

$$\Psi_{qes}(w) = \left(\left[\prod_{i=1}^{N} \left(2\ell^2 \frac{\partial}{\partial z_i} - w^* \right) \right] \phi \right) \prod_{i=1}^{N} e^{-|z_i|^2/(4\ell^2)}$$

17.3.3 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. 17.2

Fig. 17.2 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.

17.4 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 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

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

[19] Berry's work on Berry Phase in 1984 had a number of precursors, most prominentaly the work of Pancharatnam in 1956.

$$\left[-\hbar\dot{\gamma} + i\hbar\frac{\partial}{\partial t}\right]|\psi(\mathbf{R}(t))\rangle \quad = \quad E(\mathbf{R}(t))|\psi(\mathbf{R}(t))\rangle$$

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.

17.5 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. 17.2. 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

$$\Delta\gamma = -i\int_0^{2\pi} d\theta\ \langle\Psi(\theta)|\partial_\theta|\Psi(\theta)\rangle$$

where we have written $|\Psi(\theta)\rangle$ to mean $|\Psi(|w|e^{i\theta})\rangle$. We then have

$$\partial_\theta|\Psi(\theta)\rangle = \frac{\partial w}{\partial\theta}\left(\sum_i \frac{-1}{z_i - w}\right)|\Psi(\theta)\rangle$$

[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 ***

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. 17.2. 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

[22]The way this is written it is obviously a bit nonsense. Please fix it

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.

17.6 Gauge Choice and Monodromy

The Lauglin wavefunction with M quasiholes takes the form

$$\Psi(w_1,\dots,w_M;z_1,\dots,z_N)=\mathcal{N}(w_1,\dots,w_N)\left[\prod_{\alpha=1}^{M}\prod_{i=1}^{N}(z_i-w_\alpha)\right]\Psi^{(m)}_{Laughlin}(z_1,\dots,z_N) \tag{17.3}$$

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

$$|\mathcal{N}(w_1,\dots,w_M)|=C\prod_{\alpha<\beta}|w_\alpha-w_\beta|^{1/m}\prod_{\alpha=1}^{M}e^{-|w_\alpha|^2/(4\ell^{*2})}$$

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

$$\langle\Psi(w_1,\dots,w_M)|\Psi(w_1,\dots,w_M)\rangle$$

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

$$\mathcal{N}(w_1,\dots,w_N)=C\prod_{\alpha<\beta}(w_\alpha-w_\beta)^{1/m}\prod_{\alpha=1}^{M}e^{-|w_\alpha|^2/(4\ell^{*2})} \tag{17.4}$$

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

17.6.1 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

$$\frac{\partial}{\partial\tau} = \frac{\partial w}{\partial\tau}\frac{\partial}{\partial w} + \frac{\partial w^*}{\partial\tau}\frac{\partial}{\partial w^*} \tag{17.5}$$

[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.

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 anti-holomorphic 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

$$\Delta\gamma = -i\oint d\tau\langle\Psi(\tau)|\partial_\tau|\Psi(\tau)\rangle = -i\frac{1}{4\ell^{*2}}\oint(dww^* - dw^*w)$$

where we have used Eq. 17.5. 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!

17.7 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 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)

$$\delta n = j^0 = \frac{1}{2\pi m}\delta B$$

[24]The complex version of Stokes is as follows. Using $w = x + iy$

$$\int_{\partial A}(Fdw - Gdw^*) = -2i\int_A(\partial_{w^*}F + \partial_w G)dxdy$$

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

$$j^\mu = \frac{-1}{2\pi m}\epsilon^{\mu\nu\lambda}\partial_\nu \delta A_\lambda \tag{17.6}$$

We must, of course have charge conservation as well. This is easy to enforce by writing the current in the form

$$j^\mu = \frac{1}{2\pi}\epsilon^{\mu\nu\lambda}\partial_\nu a_\lambda \tag{17.7}$$

which then automatically satisfies

$$\partial_\mu j^\mu = 0$$

In this language, the effective Lagrangian that produces Eq. 17.6 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. 17.7. By differentiating the Lagrangian with respect to a_μ we generate the equations of motion Eq. 17.6.

More here

17.8 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 17.3, in the case of quasiholes) at coordinates w_α. Thus our wavefunction we write as

$$\Psi(w_1, \dots w_M; z_1, \dots z_N)$$

as written in Eq. 17.3. We then write a *pseudowavefunction* to describe some dynamics of the quasiholes which we write as

$$\phi(w_1, \dots, w_M)$$

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.

18

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. 18.1.

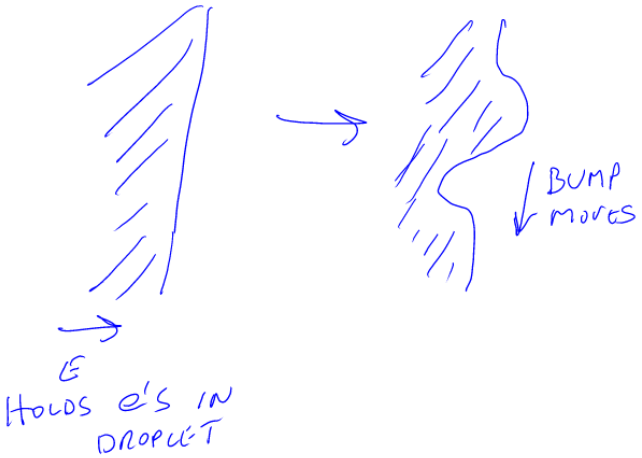


Fig. 18.1 A deformation of the edge is a low energy edge excitation which moves along the edge due to $E \times B$ drift.

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.

18.1 Landau Gauge Edge Picture for Integer Quantum Hall

(Can this section be moved to a chapter appendix?)

Recall in Landau gauge (See section 16.5) 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. 18.2.

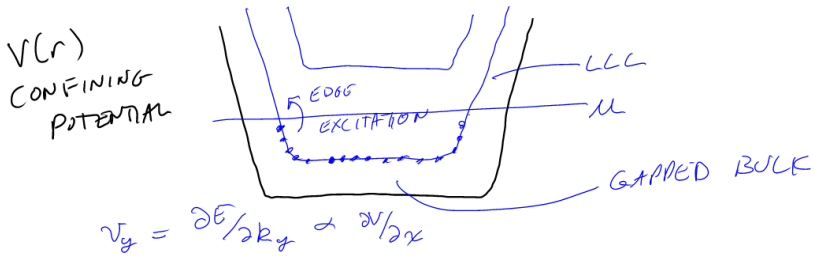


Fig. 18.2 Low energy edge excitations

The addition of the confining potential $V(x)$ simply adds this potential to the 1-d schroedinger equation 16.2. 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. 18.2.

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.

18.2 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 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

$$H_{confined} = H_0 + \alpha\hat{L}$$

[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. ($\Delta\mu = -e\Delta V$).

[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.

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.

18.3 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

$$\begin{aligned} L &= L_{ground} + \Delta L \\ E &= E_{ground} + \alpha\Delta L \end{aligned}$$

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.

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 18.1.

$L - L_{ground}$	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	$p_5, \ p_4p_1, \ p_3p_2, \ p_3p_1p_1, \ p_2p_2p_1, \ p_2p_1p_1p_1, \ p_1p_1p_1p_1p_1$	5α

Table 18.1 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])

[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 $1, 2, 3, 5, 7 \ldots$.

18.3.1 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 18.2. We see the fock space is precisely equivalent to the

$L - L_{ground}$	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 18.2 Fock Space for Chiral Bosons

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.

These operators describe a *chiral* boson – chiral because they only have positive angular momentum $m > 0$ not negative angular momentum.[6]

18.4 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).

Conformal Field Theory Approach to Fractional Quantum Hall Effect

19

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.

19.1 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 holomorphic (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

$$\sum_i \alpha_i = 0 \tag{19.1}$$

(otherwise the correlator vanishes).

[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.

19.1.1 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. 19.1 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].

19.1.2 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

$$\begin{aligned}\Psi_{qh}(w) &= \langle \psi_{qh}(w)\psi_e(z_1)\psi_e(z_2)\dots\psi_e(z_N)\hat{Q}\rangle \\ &= \left[\prod_i (z_i - w)^{\beta\sqrt{m}}\right]\Psi^{(m)}_{Laughlin}\end{aligned} \tag{19.2}$$

[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

$$e^{m\sum_{i<j}\log(z_i - z_j) - \epsilon\sqrt{m}\sum_{i,\beta}\log(z_i - z_\beta)}$$

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.

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

$$\begin{aligned} \Psi(w_1, \ldots, w_M) &= \langle \psi_{qh}(w_1) \ldots \psi_{qh}(w_M)\, \psi_e(z_1) \ldots \psi_e(z_N)\, Q \rangle \quad (19.3) \\ &= C \prod_{\alpha<\beta} (w_\alpha - w_\beta)^{1/m} \prod_{\alpha=1}^{M} e^{-|w_\alpha|^2/(4\ell^{*2})} \left[\prod_{\alpha=1}^{M} \prod_{i=1}^{N} (z_i - w_\alpha) \right] \Psi^{(m)}_{Laughlin} \end{aligned}$$

which is properly normalized

$$\langle \Psi(w_1, \ldots w_M) | \Psi(w_1, \ldots w_M) \rangle = \text{Constant}$$

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 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.

19.2 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

$$\phi_i \times \phi_j = \sum_k N_{ij}^k \phi_k$$

[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!

where fusion with the identity is trivial

$$I \times \phi_j = \phi_j$$

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

$$\lim_{w\to z} \quad \langle \psi_a(w)\psi_b(z)\ \psi_c(y_1)\psi_d(y_2)\dots\psi_n(y_m)\rangle = \sum_k C_{ab}^k (w-z)^{h_k - h_a - h_b} \langle \psi_k(z)\ \psi_c(y_1)\psi_d(y_2)\dots\psi_n(y_m)\rangle$$

19.2.1 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

$$h_{\alpha+\beta} - h_\alpha - h_\beta = \frac{(\alpha+\beta)^2}{2} - \frac{\alpha^2}{2} - \frac{\beta^2}{2} = \alpha\beta$$

19.2.2 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

$$\begin{aligned} h_I &= 0 \\ h_\sigma &= 1/16 \\ h_\psi &= 1/2 \end{aligned}$$

The fact that $h_\psi = 1/2$ is an indication that it is a fermion. The nontrivial fusion rules

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

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

$$\begin{aligned} \psi(w)\psi(z) &\sim (w-z)^{h_I - h_\psi - h_\psi} I + \dots \\ &\sim \frac{I}{w-z} + \dots \end{aligned}$$

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

$$\begin{aligned} \psi(w)\sigma(z) &\sim (w-z)^{h_\sigma - h_\sigma - h_\psi}\ \sigma(z) + \dots \\ &\sim (w-z)^{-1/2}\ \sigma(z) + \dots \end{aligned}$$

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]

$$\sigma(w)\sigma(z) \sim C^I_{\sigma\sigma}(w-z)^{-1/8} I + C^\psi_{\sigma\sigma}(w-z)^{3/8}\psi(z) + \dots \qquad (19.4)$$

[7]Remember these exponents of 1/8 and 3/8 from the ising anyon homework problems?

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

$$\lim_{w\to z}\langle\sigma(w)\sigma(z)\rangle \sim (w-z)^{-1/8} \tag{19.5}$$

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

$$\lim_{w\to z}\langle\sigma(w)\sigma(z)\psi(y)\rangle \sim (w-z)^{3/8} \tag{19.6}$$

Similarly one can see that fusion of two sigmas in the presence of any even number of ψ fields will be similar to Eq. 19.5, whereas in the presence of any odd number of ψ fields it will be like Eq. 19.6.

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

19.3 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.

[8]Insert footnote or appendix that derives this. See Yellow Book for now!

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

$$\langle \psi(z_1)\psi(z_2)\dots\psi(z_N)\rangle = \mathcal{A}\left[\frac{1}{z_1-z_2}\,\frac{1}{z_3-z_4}\dots\frac{1}{z_{N-1}-z_N}\right] \equiv \mathrm{Pf}\left(\frac{1}{z_i-z_j}\right) \tag{19.7}$$

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

$$\Psi = \mathrm{Pf}\left(\frac{1}{z_i-z_j}\right)\prod_{i<j}(z_i - z_j)^m \prod_{i=1}^{N} e^{-|z_i|^2/(4\ell^2)}$$

[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 $(\mathrm{Pf}[M])^2 = \mathrm{det}M$.

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 $\nu = 1/m$.

19.3.1 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.

[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

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.

19.4 Quasiholes of the Moore-Read state

We now try to construct quasiholes for the Moore-Read Pfaffian wavefunction. As we did in Eq. 19.2, 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.

Laughlin Quasihole

One obvious thing to try would be to write a simple vertex operator

$$\psi^L_{qh}(w) = V_\beta(w)$$

Looking at the OPE we have

$$\psi^L_{qh}(w)\psi_e(z) \sim (w-z)^{\beta\sqrt{m}}\psi(z)$$

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

$$\begin{aligned}\Psi^L_{qh}(w) &= \langle \psi^L_{qh}(w)\,\psi_e(z_1)\psi_e(z_2)\dots\psi_e(z_N)\hat{Q}\rangle \qquad (19.8)\\ &= \left[\prod_{i=1}^{N}(z_i - w)\right]\Psi^{(m)}_{Moore-Read}\end{aligned}$$

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

$$\psi_{qh}(w) = \sigma(w)V_\beta(w)$$

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.

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 19.2, 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

$$\Psi_{qh}(w,w') = \langle \psi_{qh}(w)\psi_{qh}(w')\,\psi_e(z_1)\psi_e(z_2)\ldots\psi_e(z_N)\hat{Q}\rangle \tag{19.9}$$

$$= (w-w')^{\frac{1}{4m}}\; e^{-(|w|^2+|w'|^2)/4\ell^{*2}} \prod_{i=1}^{N}(w-z_i)^{1/2}(w'-z_i)^{1/2} \tag{19.10}$$

$$\times\; \langle \sigma(w)\sigma(w')\,\psi(z_1)\psi(z_2)\ldots\psi(z_3)\rangle \prod_{i<j}(z_i-z_j)^m \prod_{i=1}^{N} e^{-|z_i|^2/(4\ell^2)}$$

Several comments are in order here. First of all, from the first line of Eq. 19.10 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 Eq. 19.7). 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.

[13]Like the Sith, they come in pairs.

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. 19.5. 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. 19.10 and the $-\frac{1}{8}$ is from the operator product expansion Eq. 19.5. 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 $h_{\text{“}I\text{”}} = 1/2m$.

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. 19.6. As the two quasiholes

[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.

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. 19.10 and the $\frac{3}{8}$ is from the operator product expansion Eq. 19.6. 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

$$[R^{\text{“}\psi\text{”}}_{qh-qh}]^2 = e^{2\pi i(\frac{1}{4m}+\frac{3}{8})}$$

19.5 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. 19.10) 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

$$G(w_1,w_2,w_3,w_4) = \langle \sigma(w_1)\sigma(w_2)\sigma(w_3)\sigma(w_4)\rangle \tag{19.11}$$

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 19.2). There are two different ways in which this can happen

$$\begin{aligned} \sigma(w_1)\sigma(w_2) &\rightarrow I \\ \sigma(w_3)\sigma(w_4) &\rightarrow I \end{aligned}$$

OR we could have

$$\begin{aligned} \sigma(w_1)\sigma(w_2) &\rightarrow \psi \\ \sigma(w_3)\sigma(w_4) &\rightarrow \psi \end{aligned}$$

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

[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.

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. 19.11. 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].

$$\lim_{w\to\infty} \langle \sigma(0)\sigma(z)\sigma(1)\sigma(w)\rangle = a_+ G_+(z) + a_- G_-(z) \tag{19.12}$$

where

$$G_\pm = (wz(1-z))^{-1/8}\sqrt{1 \pm \sqrt{1-z}} \tag{19.13}$$

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

$$\begin{aligned} \lim_{z\to 0} \quad & G_+ \sim z^{-1/8} \qquad && (\sigma(0)\sigma(z) \to I) \\ \lim_{z\to 0} \quad & G_- \sim z^{3/8} \qquad && (\sigma(0)\sigma(z) \to \psi) \end{aligned}$$

Thus (comparing to Eqs. 19.5 and 19.6) 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. 19.13 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

[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!

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 19.13 it naively looks like both

$$\begin{aligned}\lim_{z\to 1} G_+ &\sim (1-z)^{-1/8}\\ \lim_{z\to 1} G_- &\sim (1-z)^{-1/8}\end{aligned}$$

But examining this a bit more closely we realize we can construct the linear combinations

$$\begin{aligned}\tilde{G}_+ &= \frac{1}{\sqrt{2}}(G_+ + G_-)\\ \tilde{G}_- &= \frac{1}{\sqrt{2}}(G_+ - G_-)\end{aligned}$$

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.

$$\begin{pmatrix}\tilde{G}_+\\ \tilde{G}_-\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}G_+\\ G_-\end{pmatrix}$$

[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.

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. 19.1

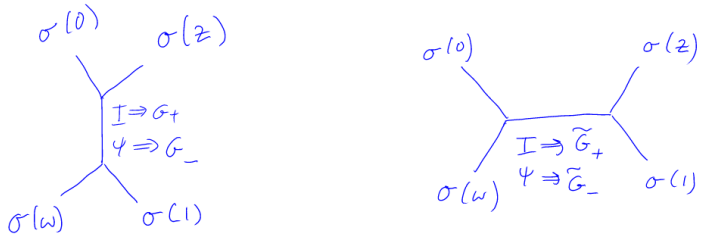


Fig. 19.1 The F-matrix transforms between the two fusion channels depicted here.

19.6 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. 19.10 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.

19.7 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 enforc-

ing 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.

19.7.1 $\mathbb{Z}_3$ 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

$$\begin{aligned} \sigma_1 &= \psi_2\tau \\ \sigma_2 &= \psi_1\tau \\ \epsilon &= \tau \end{aligned}$$

and using the Fibonacci fusions $\tau\times\tau = I+\tau$ and the $\mathbb{Z}_3$ fusions $\psi_i\times\psi_j = \psi_{(i+j)\mathrm{mod}3}$ 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

$$\psi_e(z)\psi_e(w) \sim (z-w)^m\psi_2(z)V_{2\sqrt{m+\frac{2}{3}}}(z)$$

[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 $h = 2/5+2/3 = 1+1/15$

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.

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

$$e^* = e\nu/3$$

Commentary on References

20

(1) A general reference which should be useful for much of the book is the review article by Nayak, Simon, et al [1].

(2) A wonderful little book which is really fun to read that introduces the Kauffman invariant and many other ideas of knot theory is the book Knots and Physics by Kauffman [2], now in its 3rd edition. This book really inspired me when I was a grad student. It appears to be available online in several places (not certain which, if any, are legal). Although the whole book is fun; and much of it is written at a very introductory level, mainly the end of part 1 is the most relevant part where he explains the connection of Kauffman invariant to Chern-Simons theory (and pieces get to be well beyond introductory). There is a lot in here , the deep parts are easy to gloss over.

(3) A very nice introduction to non abelian anyons and topological quantum computation is given in John Preskill's lecture notes, available online [3].

(4) Frank Wilczek has two books which both discuss Berry phase and abelian anyons [4, 5]. Both have mainly reprints in them with some commentary by Wilczek. Often it is enough to read the commentary!

(5) If you need a refresher on path integrals, consider the first 15 pages of Fabian Essler's notes [6]. Also consider the nice article by Richard MacKenzie [7]. MacKenzie includes some useful applications such as Aharanov-Bohm effect. Look mainly at the first 22 pages.

(6) The classic paper by Ed Witten which launched the field is [8]. This is a tremendously deep paper which introduces a lot of brilliant ideas. I find something new every time I read it. I find it to be tough reading in some places and easy in others.

(7) From a more mathematial viewpoint several articles by Sir Michael Atiya are very useful [9,10]. These are both introductions to topological quantum field theories. There is also a more detailed book by the same author [11]. The full book might be hard to read unless you have a very strong maths background.

(8) A rather remarkably well written and readable master's thesis (!) by Lokman Tsui on Chern- Simons theories, topological quantum field theory, and knot theory [12].

(9) There are several nice references on the structure of topological quantum field theories and diagrammatic calculus,

Parsa Bonderson's thesis: http://thesis.library.caltech.edu/2447/2/thesis.pdf
This is a more detailed version of the long article by Kitaev ("Anyons in exactly solvable models") which I mention below. Note there is some slight change of convention between the two articles.
Also a good reference is the book on Topological Quantum Computation by Zhenghan Wang
"Topological Quantum Computation", Conference Board of the Mathematical Sciences, Regional Conference Series in Mathematics, American Mathematical Society, (Providence, Rhode Island), Number 112, 2008.
If you are more mathematical, you might like the thesis of Bruce Bartlett available online here
https://arxiv.org/abs/math/0512103

(10) The monumental work "Anyons in an exactly solved model and beyond" by Alexei Kitaev, Annals of Physics 321 (2006) 2–111
available online here
https://arxiv.org/abs/cond-mat/0506438
This brings the ideas of topological quantum field theory into the condensed matter arena. This is not easy reading, but a ton of great ideas are buried in this paper.
Another work by Kitaev, "Fault-tolerant quantum computation by anyons", Annals Phys. 303 (2003) 2-30.
available online here
https://arxiv.org/abs/quant-ph/9707021
introduces the famous toric code, discusses quantum error correction, and generalizes the toric code model to arbitrary non-abelian groups.
Kitaev's work on the quantum wire (which we might get to at the end of the course) is here.
https://arxiv.org/abs/cond-mat/0010440
A brief digest of some of the many ideas introduced in these three papers is given by notes taken by Laumann of Kitaev's lectures, available here.
https://arxiv.org/abs/0904.2771
Loop gases are introduced in this paper by Freedman et al. It has a lot of sections which are hard to parse.
http://stationq.cnsi.ucsb.edu/ freedman/Publications/83.pdf
The double-fibonacci string-net is discussed in some detail in this work by Fidkowski et al,
https://arxiv.org/abs/cond-mat/0610583
The classic paper on string - nets very generally is this by Levin and Wen.
https://arxiv.org/abs/cond-mat/0404617
The standard reference on introductory quantum hall effect is the classic book, "The Quantum Hall Effect", edited by Prange and Girvin, published by Springer. The first chapter, and the chapters by Laughlin and Haldane are probably the best. The experimental chapters are good for context too.

Another decent reference quantum Hall physics is T. Chakraborty and P. Piettilainen, "The Quantum Hall Effects: Integral and Fractional," (Springer 1995).
A short review article by Macdonald is pretty nice and is available here.
https://arxiv.org/pdf/cond-mat/9410047v1.pdf
The article that introduced the ideas of conformal field theory into the field of quantum Hall effect is by Moore and Read, available online here.
http://www.physics.rutgers.edu/ gmoore/MooreReadNonabelions.pdf
A recent review article on Fractional quantum Hall hierarchies (and also discusses nonabelian quantum Hall and conformal field theory) is online here.
https://arxiv.org/abs/1601.01697

A few random digressions:

(11) If you are interested in 2+1 D quantum gravity, see this article . I can't vouch for it, but the introduction is interesting;
https://link.springer.com/article/10.12942/lrr-2005-1
This is the article by Witten explaining how 2+1 D gravity is "exactly solvable." More from Witten here. There is reconsideration many years later, again by Witten, see here .
http://www.sciencedirect.com/science/article/pii/0550321389905919

(12) I've been told the book by Jiannis Pachos on topological quantum computation is a good resource.

(13) If you are interested in the topology of manifolds in 3 and 4 dimensions, there are several good books. One by Kirby is online here.
https://math.berkeley.edu/ kirby/papers/Kirby
There is a book by Gompf and Stipcitz "4-manifolds and Kirby Calculus" which is nice. Note that parts of this book are online free if you google them.
https://www.amazon.co.uk/4-Manifolds-Calculus-Graduate-Studies-Mathematics/dp/0821809946

(14) For more information on conformal field theory. The standard reference is the Big yellow book (Conformal Field Theory Authors: Philippe Di Francesco, Pierre Mathieu, David Snchal) . The first part of this book (up to chapter 12) is excellent, but even that much is a lot of reading. There is a short set of lectures from les Houches by Ginsparg .
https://arxiv.org/abs/hep-th/9108028
I also like the short set of notes by Fendley .
http://galileo.phys.virginia.edu/ pf7a/msmCFT.pdf
For even shorter introduction of what you need to apply CFT to quantum Hall, see the appendix of Ref. 1 above, or the appendix of ***.
The book by Kauffman and Lins gives more details of constructing a full anyon theory from the kauffman invariant.

http://press.princeton.edu/titles/5528.html
Neilsen and Chuang for quantum computation in general, although there are plenty of other refs.

References

[1] *Non Abelian Anyons and Topological Quantum Computation*, Chetan Nayak, Steven H. Simon, Ady Stern, Michael Freedman, Sankar Das Sarma; Rev. Mod. Phys. 80, 1083 (2008). Also available online at https://arxiv.org/abs/0707.1889

[2] *Knots and Physics*, by Louis Kauffman, World Scientific, (2001), 3ed.

[3] *Lecture Notes on Topological Quantum Computation*, John Preskill, http://www.theory.caltech.edu/ preskill/ph219/topological.pdf

[4] *Geometric Phases in Physics* F. Wilczek and A. Shapere eds, World Scientific, (1989).

[5] *Fractional Statistics and Anyon Superconductivity*, F. Wilczek ed, World Scientific, (1990).

[6] *Lecture Notes for the C6 theory option*, F. Essler, http://www-thphys.physics.ox.ac.uk/people/FabianEssler/C6web2012/lecturenotes2015.pdf

[7] *Path Integral Methods and Applications*, R. MacKenzie, https://arxiv.org/abs/quant-ph/0004090

[8] *Quantum Field Theory and the Jones Polynomial* E. Witten, Comm. Math. Phys. Volume 121, Number 3 (1989), 351-399; available online here https://projecteuclid.org/euclid.cmp/1104178138

[9] *Proceedings of 5th Gokova Geometry and Topology Conference*, M. F. Atiyah, Tr. J. Mathematics, 21, 1, (1997). http://www.maths.ed.ac.uk/ aar/papers/atiyahinttqft.pdf

[10] *Topological quantum field theory.* M. F. Atiyah, Publications Mathmatiques de l'IHS, 68 (1988), p. 175-186 http://www.numdam.org/item?id=PMIHES_1988__68__175_0

[11] *The Geometry and Physics of Knots*, M. F. Atiyah, ,Cambridge U. Press, 1990.

[12] *Application of Chern-Simons theory in Knot Theory*, Master's Thesis, Lokman Tsui, http://ms.mcmaster.ca/∼boden/students/Tsui-MSc.pdf