Topological Quantum: Lecture Notes and Proto-Book

Steven H. Simon

Comments About This draft

This is a set of course notes hoping to someday be a book.

Unfortunately, there is a huge difference between course notes and a book. This is why I need everyone's help. If there are parts that you think are unclear – please let me know. If there are errors — please let me know (even if they are small and subtle). If the figures are unclear — please let me know. If there are mistakes in grammar — please let me know.

If you don't get the jokes... well, that is your problem. Seriously though, I need help if this is eventually going to arrive at the Nirvana that is bookdom. Give me feedback please. It will be good for your Karma. ☺

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, 2018, 2019, 2020, ... and this kept going until I finished the book, which seemed like forever.

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. In the course I teach, I am certainly not assigning all of the chapters — I'm not a sadist!

I should also mention that chapter 33 introduces some basic mathematics that many people may know but I thought should be included for completeness.

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. Many technical details are pushed to chapter appendices — often these can be skipped on a first reading.

In a margin note of my previous book, I said that my next book (i.e., this one) would be about two dimensional electron systems. This topic is covered in the section on fractional quantum Hall effect[1].

A list of useful references is given etc.

[1] I also suggested that I might write a thriller about physicists defeating drug smugglers. For those who are interested, I'm still working on it, but I discovered that writing a novel is pretty hard.

Contents

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

1

The field of quantum topology 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 precedents 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 chapters 28 and 27 we will discuss some of the key works that this Nobel Prize honored.

[2] We will see this starting in chapter 24 below.

[3] See chapter 6.

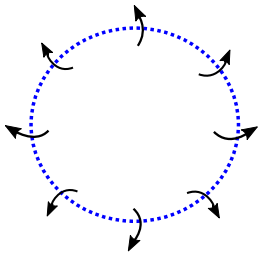


Fig. 1.1 A smoke ring or vortex loop is an invisible ring in space where the fluid flows around the invisible ring as shown by the arrows. The whole thing moves out of the plane of the page at you as the fluid circulates.

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

[5] Even in 1867, a talented smoker could produce a smoke ring from their mouth.

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

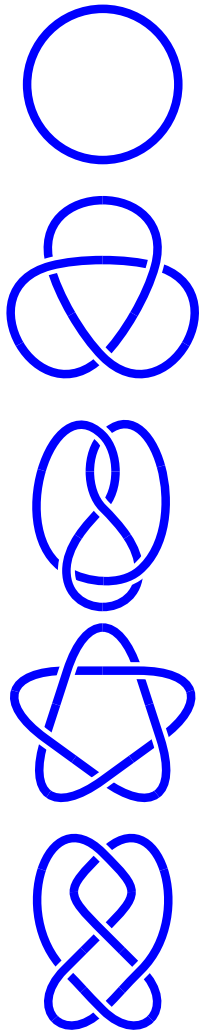


Fig. 1.2 The simplest few knots made from one strand of string. The top knot, a simple loop, is known as the "unknot", and corresponds to the simple smoke ring in Fig. 1.1. The second knot from the top, known as the trefoil, is not the same as its mirror image (see exercise 2.1)

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 one 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 shown 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 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 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 understanding atoms was probably unrelated to understanding knots. Tait died[8] in 1901 not realizing that his work on the theory of knots 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 Stoimenow [2008] 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.

Further Reading

- Daniel S. Silver, "Knot Theory's Odd Origins", American Scientist, Volume 94, 2006.

Kauffman Bracket 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.

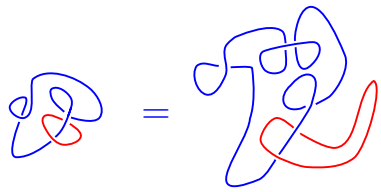


Fig. 2.1 Topological equivalence of two knots. The knot on the left can be deformed continuously into the knot on the right without cutting any strands.

It may appear easy to determine whether two simple knots are topologically equivalent and when they are not. However, for complicated knots, it becomes *extremely difficult* to determine whether two knots are equivalent or inequivalent. It is thus useful to introduce a mathematical tool known as a knot invariant that can help us establish when two knots are topologically inequivalent.

A **Knot Invariant** is a mapping from a knot (or a picture of a knot) to an output via a set of rules which are cooked up in such a way that two topologically equivalent knots must give the same output. (See Fig. 2.2.) So if we put two knots into the set of rules and we get two different outputs, we know immediately that the two knots cannot be continuously deformed into each other without cutting.

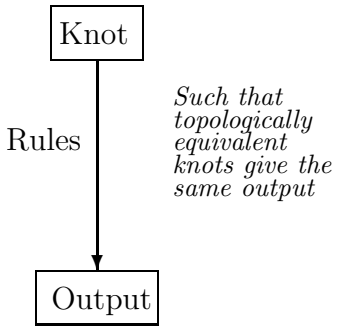


Fig. 2.2 Schematic description of a knot invariant as a set of rules taking an input knot to some mathematical output such that topologically equivalent knots give the same output.

To demonstrate how knot invariants work, we will use the example of the Kauffman bracket invariant[2,3] (See Kauffman [1987]). The Kauff-

[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. In either case no dangling ends are allowed. 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 2.6.1). Two knots are isotopic if there is a continuous smooth family of knots between the initial knot and the final knot — however to be more precise, as we will see below in section 2.2.1, we should think of the knots as being thickened to ribbons and we want a smooth family of ribbons.

[2] Be warned: there are multiple things named after Kauffman. The particular normalization of the bracket invariant that we use has been named the *topological bracket* by Kauffman. The more common definition of the bracket is our definition divided by d.

[3] The term "bracket" is due to a common notation where one draws a picture of a knot inside brackets to indicate that one is supposed to evaluate this invariant. We will not draw these brackets.

man bracket invariant was essentially invented by Vaughn Jones who won the Fields medal for his work on knot theory[Jones, 1985]. Kauffman's important contribution to this story (among his many other contributions in the field of knot theory) was to explain Jones' work in very simple terms.

To define the **Kauffman Bracket Invariant**, we start with a scalar variable A. For now, leave it just a variable, although later we may give it a value. There are then just two rules to the Kauffman bracket invariant. First, a simple loop of string (with nothing going through it) can be removed from the diagram and replaced with the number[4]

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

[4]We will eventually see that d stands for "dimension".

The second rule replaces a diagram that has 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 as shown in Fig. 2.3. This type of replacement rule is known as a *skein* rule.[5]

[5]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 this is unrelated.

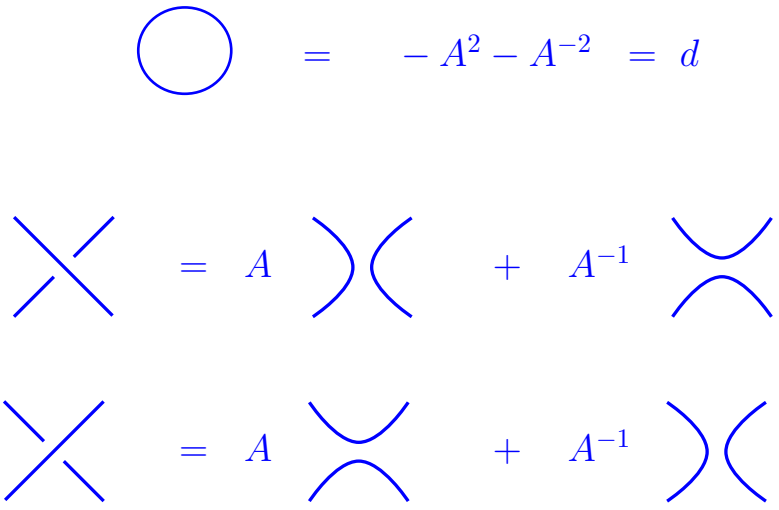


Fig. 2.3 Rules for evaluating the Kauffman bracket invariant. The third line is exactly the same as the middle line except that all the diagrams are rotated by 90 degrees, so it is not an independent rule. However, it is convenient to draw the rule twice to make it easier to compare to other diagrams.

The general scheme is to use the second (and third) 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.

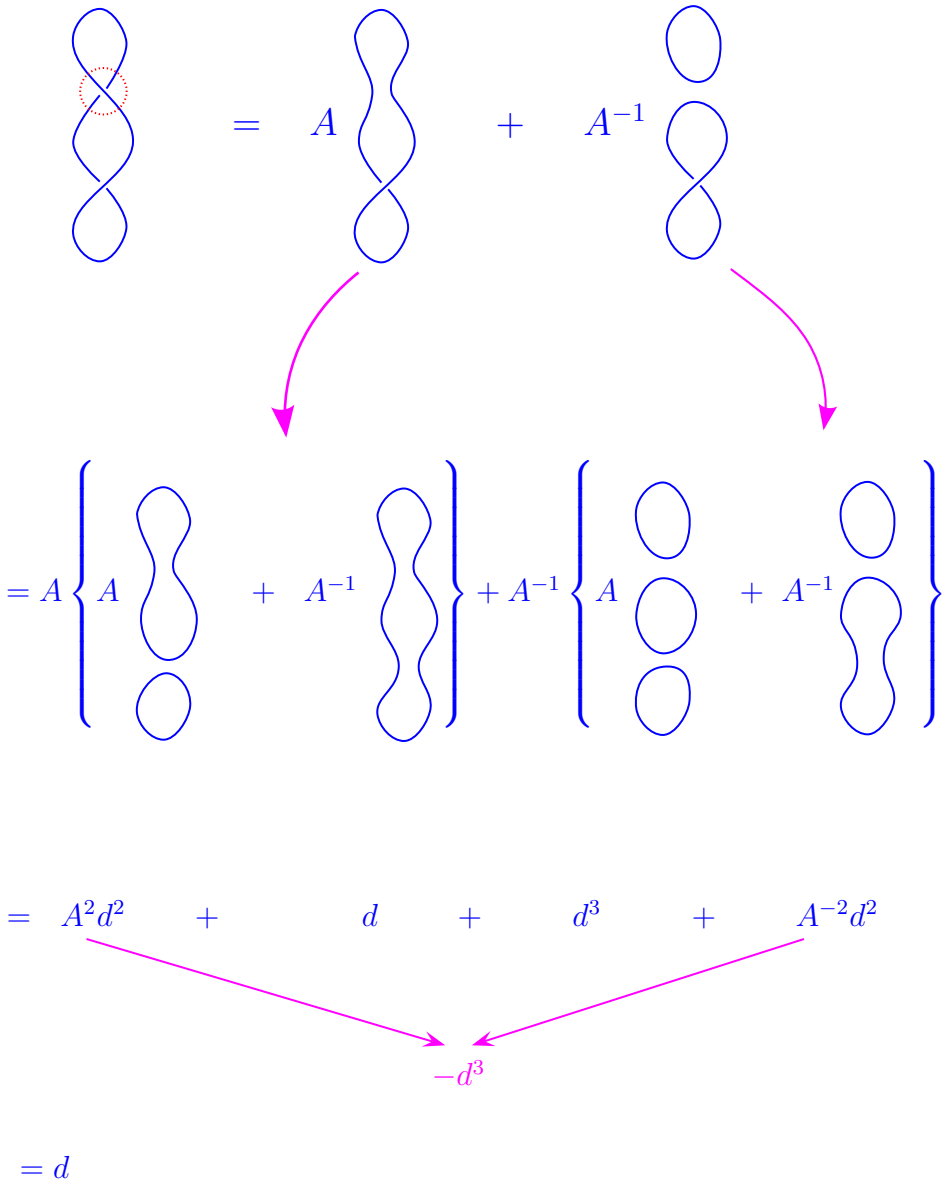


Fig. 2.4 Example of evaluation of the Kauffman bracket invariant for the simple twisted loop in the upper left. The light dotted red circle is meant to draw attention to where we apply the Kauffman crossing rule (the middle line in Fig. 2.3) to get the two diagrams on the right hand side. After applying the Kauffman rules again (the final line in Fig. 2.3), we have removed all crossings and 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.

To give an example of how these rules work we show evaluation of the Kauffman bracket invariant for the simple knot in the upper left of Fig. 2.4. The output of the calculation is that the Kauffman invariant of this knot comes out to be d. This result 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[6] 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[7]. 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 Kauffman bracket invariant of this knot! So one might then think that this Kauffman bracket invariant is actually not so useful for complicated knots. We will return to this issue later in Section 2.4.

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

[7] 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 (a simple loop) which has Kauffman invariant d. It is also an open challenge to find out whether any combinatoric knot invariants similar to Kauffman can distinguish all topologically inequivalent knots from each other.

2.2 Relation to Physics

There is a fascinating relationship between knot invariants and quantum physics. For certain types of so-called "topological quantum systems" the amplitudes of space-time processes can be directly calculated via knot invariants such as the Kauffman bracket 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.[8]

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.

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

[8] There is also some discussion of "topological" systems in 1+1 D in chapter 12 for example.

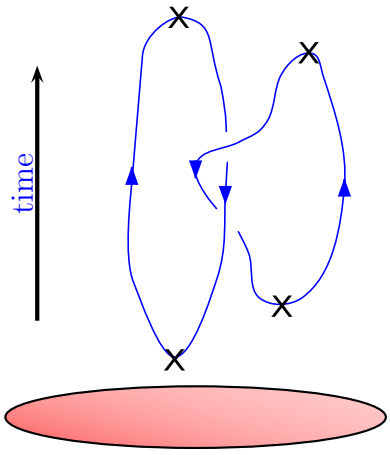


Fig. 2.5 A space-time process showing world lines of particles for a 2+1 dimensional system (shown as the shaded disk at the bottom). The X's mark the points in space-time where particles-anti-particle pairs are either pair-created or pair-annihilated.

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 Fig. 2.5. It should surprise us that systems exist where amplitudes depend only on topology, 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 [Witten, 1989]. He won the Fields medal along with Vaughan Jones for this work.

Such topological theories were first considered as an abstract possibility, mainly coming from researchers in quantum gravity (see chapter 6). However, now several systems are known in condensed matter which actually behave like this. While not all topological theories are related to the Kauffman bracket invariant, many of them are (There are other knot invariants that occur in physical systems as well — including the so-called HOMFLY invariant[Freyd et al., 1985]. See exercise 29.) 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.

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

(1) $SU(2)_2$ **class**. For these, the Kauffman bracket invariant gives the quantum amplitude of a process by using the value $A = ie^{-i\pi/(2(2+2))} = i^{3/4}$. This is also known as "Ising" anyons[9]. Possibly physical realizations include

- $\nu = 5/2$ Fractional Quantum Hall Effect (2D electrons at low temperature in high magnetic field). See chapters ***.
- 2D p-wave superconductors.
- 2D Films of ^{3}HeA superfluid[10].
- 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. See chapter ***?

(2) $SU(2)_3$ **class**. For this, the Kauffman bracket invariant gives the quantum amplitude of a process by using the 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) $SU(2)_4$ **class**. For this, the Kauffman bracket invariant gives the quantum amplitude of a process by using the 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) $SU(2)_1$ **class** Also known as semions. These are proposed to be realized in rotating boson fractional quantum Hall effect (See comments in chapter 31). This corresponds to a fairly trivial knot invariant as we will see later in section ***.

(5) $SU(3)_2$ **class**. This corresponds to a case of the HOMFLY knot invariant rather than the Kauffman bracket invariant. It is possible that the unpolarized $\nu = 4/7$ fractional quantum hall effect is in this class.

[9] 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 chapter 31.

[10] 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 5). The slight differences are related to extra phases that appear in braiding. See also chapter ****. See end of chapter for references ***

2.2.1 Twist and Spin-Statistics

Before moving on, let us do some more careful examination of the Kauffman bracket invariant. To this end, let us examine a small loop in a piece of string (as shown in Fig. 2.6) and try to evaluate its Kauffman bracket invariant.

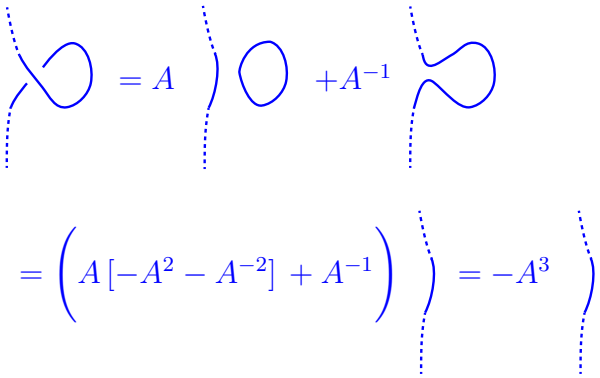


Fig. 2.6 Evaluation of a twist loop in a string. The dotted lines going off the top and bottom of the diagrams mean that the string will be connected up with itself, but we are not concerned with any part of the knot except for piece shown. The result of this calculation is that removal of the little 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 bracket 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[11]. 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"[12]. 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 tried to straighten a garden hose will realize this![13]

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.

[11] In mathematics we say they are ambient isotopic but not regular isotopic! (See section 2.6.1)

[12] 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 a knot as a line, one must have a convention as to what this means for the orientation of the ribbon. See comment on blackboard framing at the end of this section.

[13] If you have not had this experience with a garden hose, you are not paying enough attention to your garden!

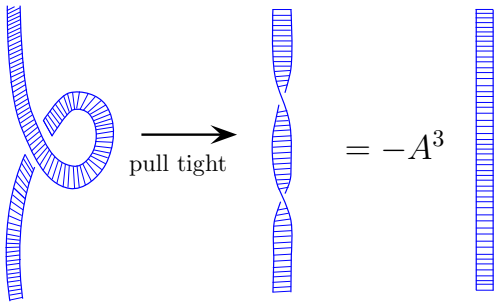


Fig. 2.7 Pulling straight a small loop introduces a twist in the string. This twist can be replaced with a factor of $-A^3$.

In fact, Fig. 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 3, in 2+1 D 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[14].

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 bracket invariant and removes a factor of $-A^3$ for each self twist that occurs[15]. This gives the famed Jones Polynomial knot invariant. See exercise 2.4.

[14] 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 3.

[15] To properly count the self twists, one calculates the so-called "writhe" of the knot (See section 2.6.2). 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 positive 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.

Blackboard Framing

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. 2.8. If we intend a strand to have a self twist, we draw it as a loop as in the left of Fig. 2.7 or the left of Fig. 2.6.

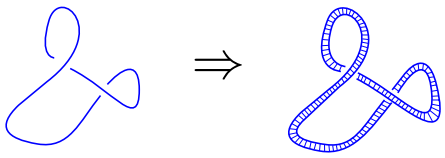


Fig. 2.8 Blackboard framing. The knot drawn on the left represents the ribbon on the right, where the ribbon always lies flat in the plane of the page (i.e., the plane of the blackboard).

2.3 Bras and Kets

For many topological theories (the so-called nonabelian theories) the physical systems have an interesting, and very unusual property. Imagine we start in a ground state (or vacuum) 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 Fig. 2.9.

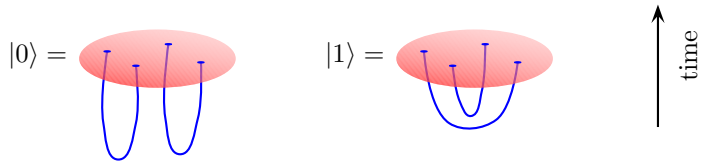


Fig. 2.9 Two linearly independent quantum states that look identical locally but have different space-time history. The horizontal plane is a space-time slice at fixed time, and the diagrams are all oriented so time runs vertically.

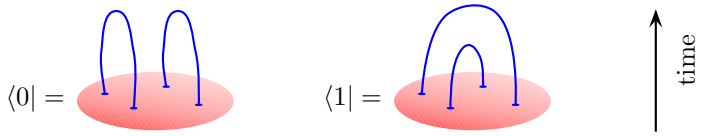


Fig. 2.10 Kets are turned into bras by reversing time.

To demonstrate that these two different space-time histories are linearly independent quantum states, we simply take inner products as shown in Fig. 2.11 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 $1/d$ in order that we have normalized states.)

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. So for example, the inner product $\langle 0|1\rangle$ as shown in the bottom of Fig. 2.11 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.

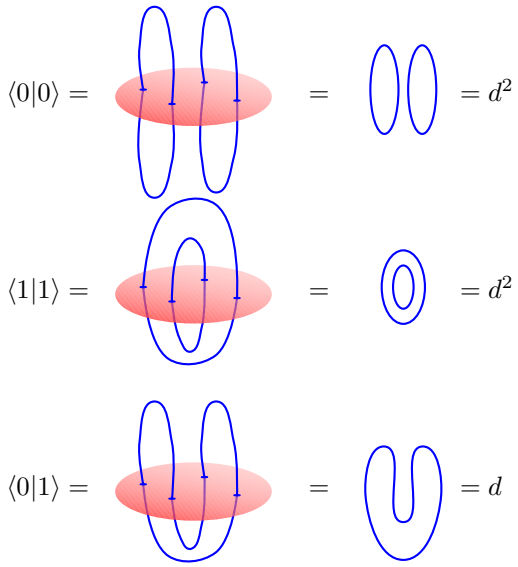


Fig. 2.11 Showing that the kets $|0\rangle$ and $|1\rangle$ are linearly independent. For $|d| \neq 1$ the inner products show they must be linearly independent quantities.

Suppose now we insert a braid between the bra and the ket as shown in Fig. 2.12. The braid makes a unitary operation on the two dimensional vector space spanned by $|0\rangle$ and $|1\rangle$. We can once again evaluate this matrix element by calculating the Kauffman bracket invariant of the resulting knot.

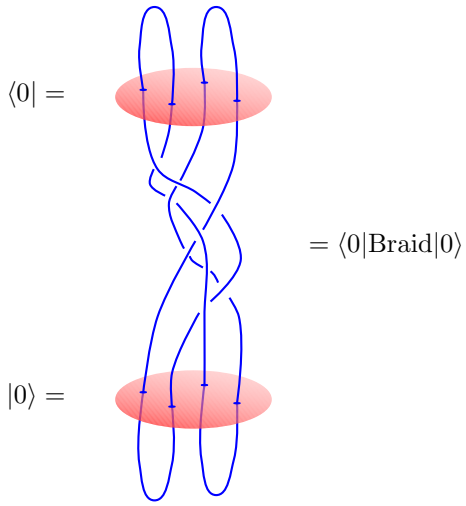


Fig. 2.12 Inserting a braid between the bra and the ket. The braid performs a unitary operation on the two dimensional vector space spanned by $|0\rangle$ and $|1\rangle$

2.4 Quantum Computation with Knots

Why do we care so much about topological systems and knot invariants? A hint is from the fact that we wrote states above as $|0\rangle$ and $|1\rangle$. This notation suggests the idea of qubits[16], and indeed this is one very good reason to be interested.

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 bracket invariant[17]. 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. Some of the particles will reannihilate, and others will refuse to go back to the vacuum (forming bound states instead). The probability that they all reannihilate is (up to a normalization[18]) given by the absolute square of the Kauffman bracket invariant of the knot (since amplitudes are the Kauffman bracket invariant, the square of the Kauffman bracket invaraint is the probability). Even estimation of the Kauffman bracket 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![19] Thus the topological quantum system has computational ability beyond that of a classical computer.

It turns out that the ability to calculate Kauffman bracket invariant is sufficient to be able to do any **quantum computation**[20]. One can use this so-called **topological quantum computer** to run algoritms such as Shor's famous factoring (i.e., code breaking) algorithm[21]. The idea of using topological systems for quantum computation is due to Michael Freedman and Alexei Kitaev[22].

So it turns out that these topological systems can do quantum computation. Why is this a good way to do quantum computation?[16]. 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

[16] 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! We will discuss quantum computation more in chapter 11.

[18] 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 bracket 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 bracket invariant.

[19] The details of this are a bit subtle and are discussed by Aharonov et al. [2009]; Aharonov and Arad [2011]; Kuperberg [2015].

[20] In fact the computational power of being able to evaluate the Kauffman bracket for fixed A is equivalent to the computational power of a quantum computer, with the exception of a few special values of the Kauffman parameter A.

[21] See Nielsen and Chuang [2000], for example, for more detail about quantum computation in general.

[22] 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. Freedman is also a champion rock climber.

[17] 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 following year Feynman started 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.

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. 2.13. 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.)

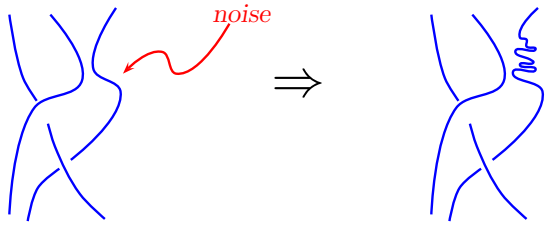


Fig. 2.13 The effect of noise on a topological quantum computation. As long as the noise does not change the topology of the knot, then no error is introduced.

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.

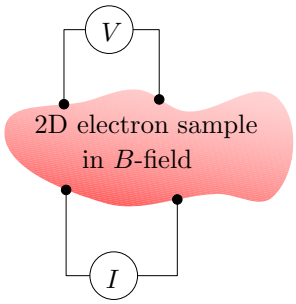
FQHE occurs in two dimensional electronic systems[23] in high magnetic field at low temperature (typically below 1K). There are many FQHE states which are labeled by their so called filling fraction $\nu = p/q$ with p and q small integers. The filling fraction can be changed in experiment by, for example, varying the applied magnetic field (we will discuss this later in chapter **??**). The FQHE state emerges at low temperature and is topological[24].

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 resistance measurement, as shown in Figs. 2.14 and 2.15. In , Fig. 2.14 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.

The measurement in the Fig. 2.15 is more interesting. In this case, the Hall voltage is precisely quantized as $V = (h/e^2)(1/\nu)I$ where I is

[23] Electronic systems can be made two dimensional in several ways. See comments in chapter **??**.

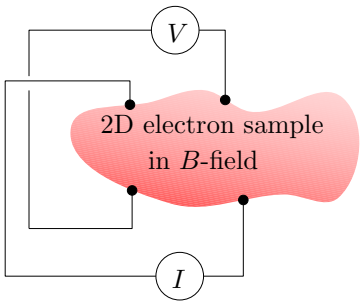
[24] A comment in comparing this paradigm to the common paradigm of high energy physics: In high energy there is generally the idea that there is some grand unified theory (GUT) at very high energy scale and it is extremely symmetric, but then when the universe cools to low temperature, symmetry breaks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



$V = 0$

Dissipationless Flow

Fig. 2.14 Measurement of longitudinal resistance in FQHE experiment.



$V = (h/e^2)(1/\nu)I$

Quantized Resistance

Fig. 2.15 Measurement of Hall resistance in FQHE experiment.

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. 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 Fig. 2.15, but not in Fig. 2.14. We should emphasize that this is extremely unusual. If you were to measure the resistance of a bar of copper, the voltage would depend entirely on how far apart you put 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. As we might expect for a dissipationless fluid, the vortices are persistent — they will last forever unless annihilated by antivortices.

So in fact, Kelvin was almost right (See chapter 1). 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 space-time knots!

2.6 Appendix: More Knot Theory Basics

2.6.1 Isotopy and Reidermeister Moves

Two knots (or two pictures of knots) are *ambient isotopic* if one can be deformed into each other other without cutting any of the strands. 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[25], as shown in Fig. 2.16.

In the context of quantum physics, and as elaborated in section 2.2.1, we are usually concerned with *regular isotopy* which treats the strands as ribbons. Two knots are regular isotopic if they can related to each other using only type-II and type-III moves. A type-I move inserts a twist in the ribbon (See Fig. 2.7) and results in a different ribbon diagram, whereas type-II and III moves do not twist the ribbon[26].

[25]This is a very old result, by Kurt Reidemeister from 1927. 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, 2015].

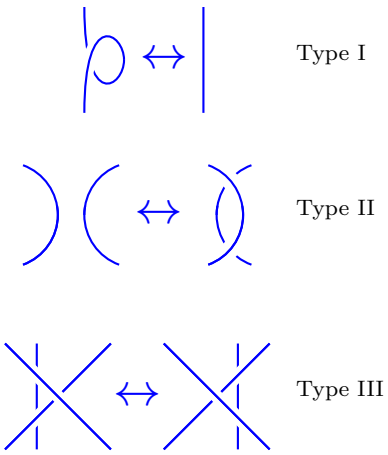
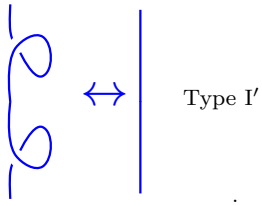


Fig. 2.16 The Three Reidermeister Moves. Any two knots that can be deformed into each other without cutting (they are "ambient isotopic") can be connected by a series of Reidermeister moves. Strictly speaking the Reidermeister moves includes the moves drawn here as well as the front-back mirror-reflections of each of these moves (turn all over-crossings to under-crossings).

[26]For regular isotopy of link diagrams one should allow cancellation of opposite ribbon twists which is sometimes known as a type-I′ move.



.

2.6.2 Writhe and Linking

Let us put arrows on all strands of our knots and links (so now we have directed lines). For each crossing we define a sign ϵ as shown in Fig. 2.17

Fig. 2.17 Defining a sign $\epsilon = \pm 1$ for each crossing of oriented knots and links.

The **writhe** w of an oriented knot (here "knot" means made of a single strand) is the sum of all of the ϵ values of the crossings

$$w(knot) = \sum_{\text{crossings}} \epsilon(\text{crossing}) \tag{2.2}$$

Note that type II and III Reidermeister moves preserve the writhe of a knot, whereas type I moves do not. Thus, the writhe is an invariant of regular isotopy but not of ambient isotopy.

For a link made of two strands, the **linking number** lk between the two strands is given by

$$lk(link) = \sum_{\substack{\text{crossings between} \\ \text{two different strands}}} \epsilon(\text{crossing}) \tag{2.3}$$

Chapter Summary

- Knot invariants, such as the Kauffman bracket 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

- The book by Kauffman [2001] is a delightful introduction to knot theory and connections to physics. This was the book that got me interested in the subject back when I was in grad school and changed the course of my life.
- I wrote another easy reading introduction, Simon [2010], connecting knots to anyons.
- Some nice introductory books on knots include Adams [1994], and Sossinsky [2002].

Exercises

Exercise 2.1 Trefoil Knot and the Kauffman Bracket

Using the Kauffman rules, calculate the Kauffman bracket invariant of the right and left handed trefoil[27] knots shown in Fig. 2.18. Conclude these two knots are topologically inequivalent. While this statement appears obvious on sight, it was not proved mathematically until 1914 (by Max Dehn). It is trivial using this technique!

Exercise 2.2 Abelian Kauffman Anyons

Anyons described by the Kauffman bracket invariant with certain special values of the constant A are abelian anyons – meaning that an exchange introduces only a simple phase as shown in Fig. 2.19.

(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., $e^{i\theta} = \pm 1$).

[27] The word "trefoil" is from the plant trifolium, or clover, which has compound trifoliate leaves.

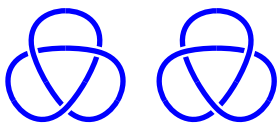


Fig. 2.18 Left and Right Handed Trefoil Knots (on the left and right respectively)

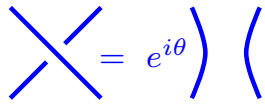


Fig. 2.19 For abelian anyons, exchange gives a phase $e^{i\theta}$.

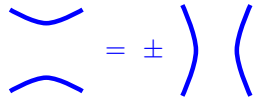


Fig. 2.20 For bosons or fermions the sign in this figure is +, for semions the sign is −.

(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 this book (See section ***). This particular phase of quantum Hall matter has been produced experimentallyClark et al. [2020], but only in very small puddles so far and it has not been possible to measure braiding statistics as of yet.

HINT: For (a) and (b) show first the identity shown in Fig. 2.20.

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

Exercise 2.3 Reidermeister moves and the Kauffman Bracket

Show that the Kauffman bracket invariant is unchanged under application of Reidermeister move of type II and type III. Thus conclude that the Kauffman invariant is an invariant of regular isotopy.

Exercise 2.4 Jones polynomial

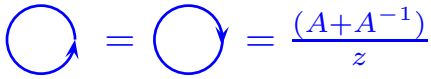
Let us define the Jones polynomial of an oriented knot as

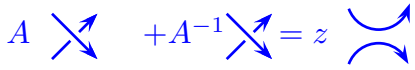
$$\text{Jones(knot)} = (-A^3)^{w(\text{knot})}\,\text{Kauffman(knot)}$$

where w is the writhe (We must first orient the knot, meaning we arrows on the strands, in order to define a writhe). Show that this quantity is an invariant of ambient isotopy – that is, it is invarient under all three Reidermeister moves.

Exercise 2.5 HOMFLY Polynomial

The HOMFLY[28] polynomial is a generalization of the Jones polynomial which has two variables A and z rather than just one variable. To define the HOMFLY polynomial we must first orient the strings in our knot or link (meaning we put arrows on the lines). The HOMFLY polynomial (Freyd et al. [1985]) of an oriented link is then defined in terms of two variables A and z by the two rules





(a) Show that the HOMFLY polynomial is invariant under type I Reidermeister moves[29]

(b) Show that the HOMFLY polynomial is invariant under type II and III Reidermeister moves, showing that it is an invariant of ambient isotopy.

[28] HOMFLY is an acronym of the names of the inventors of this polynomial. Sometimes credit is even more distributed and it is called HOMFLYPT.

[29] In order for the HOMFLY invariant to represent quantum particles, a phase for this type I move must be inserted by hand.

Part I

Anyons and Topological Quantum Field Theories

Particle Quantum Statistics

3

In chapter 2 we discussed braiding particles around each other, or exchanging their positions. This is often what we call particle statistics (or quantum statistics, or exchange 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]

[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] 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 3.2. 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! See the Further Reading for this chapter for a better introduction.

3.1 Single Particle Path Integral

Consider a space-time trajectory of a single 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.

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 $\psi(x) = \langle \mathbf{x} | \psi \rangle$ 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 composition: 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

$$\mathbf{1} = \int d\mathbf{x}_m |\mathbf{x}_m\rangle\langle\mathbf{x}_m|.$$

Feynman's genius was to realize that you can subdivide time into infinitesimally 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

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

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. 3.1 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

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

Analogous to when we evaluate regular integrals of things that look like $\int dx \, e^{iS[x]/\hbar}$, we can approximate the value of this integral in the small $\hbar$, or classical, limit by saddle point approximation. We do this by looking for a minimum of S with respect to its argument — this is where the exponent 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!

3.2 Two Identical Particles

We would now like to generalize the idea of a path integral to systems with multiple identical particles, starting with the case of two particles. If the particles are identical 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. Instead, we can only say that there are particles at both positions $\mathbf{x}_1$ and $\mathbf{x}_2$. To avoid the appearance of two different states expressed as $|\mathbf{x}_1, \mathbf{x}_2\rangle$ versus $|\mathbf{x}_2, \mathbf{x}_1\rangle$ (which are actually the same physical state![4]), it is then useful to simply agree on some convention for which coordinate we will always write first — for example, maybe we always write the leftmost particle first[5]. For simplicity, we can assume that $\mathbf{x}_1 \neq \mathbf{x}_2$, i.e., the particles have hard cores and cannot overlap[6]. For these indistinguishable particles, the Hilbert space is then cut in half compared to the case of two *distinguishable* particles where $|\mathbf{x}_1, \mathbf{x}_2\rangle$ and $|\mathbf{x}_2, \mathbf{x}_1\rangle$ mean physically different things.

We call the space of all states the configuration space $\mathcal{C}$. 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.

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

We will call the non-exchange path TYPE +1 (left in Fig. 3.1), and the exchange path TYPE −1 (right in Fig. 3.1). 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 under-crossings, 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 first, then follow the second. We can write a multiplication table for such composition of paths (the path types form a *group*, see Section 33.2)

$$\begin{array}{lllll} \text{TYPE } +1 & \text{Followed by} & \text{TYPE } +1 & = & \text{TYPE } +1 \\ \text{TYPE } +1 & \text{Followed by} & \text{TYPE } -1 & = & \text{TYPE } -1 \\ \text{TYPE } -1 & \text{Followed by} & \text{TYPE } +1 & = & \text{TYPE } -1 \\ \text{TYPE } -1 & \text{Followed by} & \text{TYPE } -1 & = & \text{TYPE } +1 \end{array} \tag{3.3}$$

So for example, an exchange path (which switches the two particles) followed by another exchange path (which switches again) results in a

[4] Often books define $|\mathbf{x}_1, \mathbf{x}_2\rangle = -|\mathbf{x}_2, \mathbf{x}_1\rangle$ for fermions. The two kets describe the same state in the Hilbert space only with a different phase prefactor. We should contrast this to the case of distinguishable particles where $|\mathbf{x}_1, \mathbf{x}_2\rangle$ and $|\mathbf{x}_2, \mathbf{x}_1\rangle$ have no overlap for $\mathbf{x}_1 \neq \mathbf{x}_2$

[5] This ordering scheme works in one dimension. In two dimensions we would perhaps say, the particle with the smaller x coordinate is written first, but in case of two particles with the same value of x, the particle with smaller y coordinate is written first.

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

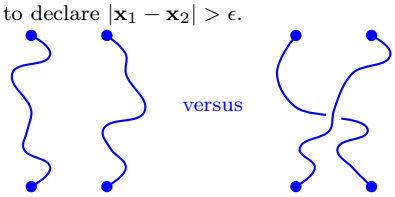


Fig. 3.1 Two possible sets of paths (paths in configuration space) from the same two starting positions to the same two ending positions (we are implying that time runs vertically). We call the non-exchange path TYPE +1, and the exchange path TYPE −1. Here we mean that time runs vertically. The two sets 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 under-crossings, but for now we will only be concerned with exchanges (TYPE −1) and non-exchanges (TYPE +1)

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 separate sums over the two different classes of paths.

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

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

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

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

$$\langle \mathbf{x}_{1f}\mathbf{x}_{2f}|\hat{U}(t_f,t_i)|\mathbf{x}_{1i}\mathbf{x}_{2i}\rangle = \tag{3.4}$$

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

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, let us check to see if the composition law is still obeyed. Again, we break the time propagation at some intermediate time

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

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

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

where in the last line we have substituted in Eq. 3.4 for each of the two propagators on the right, and we have used a bit of shorthand in writing the result.

Now, when we compose together subpaths from $i \to m$ with those from $m \to f$ to get the overall path, the sub-path types multiply according to our above multiplication table Eq. 3.3. For the full path, there are two ways to obtain a TYPE $+1$ path: (1) both sub-paths are TYPE $+1$ or (2) both sub-paths are TYPE -1. In either case, note that the net prefactor of the overall TYPE $+1$ path is $+1$. (In the case where both

subpaths are of TYPE -1, the two prefactors of -1 cancel each other). Similarly, we can consider full paths with overall TYPE -1. In this case, exactly one of the two sub-paths must be of TYPE -1, in which case, the overall sign ends up being -1. Thus, for the full path, we obtain exactly the intended form written in Eq. 3.4. I.e., under composition of paths, we preserve the rule that TYPE $+1$ paths get a $+1$ sign and TYPE -1 paths get a -1 sign. Thus this is consistent for quantum mechanics, and indeed, this is exactly what happens in the case of fermions.

3.3 Many Identical Particles

Generalizing this idea, 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.

Let us first discuss our configuration space. If we had N *distinguishable* particles in D dimensions we would have a configuration space $(\mathbb{R}^D)^N$ representing the coordinates $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_N\}$. For simplicity we usually assume all of these coordinates are different (We might imagine that the particles are hard spheres of some very small diameter ϵ). Thus we write the configuration space as $[(\mathbb{R}^D)^N - \Delta]$ where Δ represents the so-called *coincidences* where two particles are at the same position[7].

In the case of *identical* particles we want to disregard the order in which we write the coordinates. In other words, we have an equivalence relationship $\sim$ between the $N!$ possible orderings of the coordinates

$$\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_N\} \sim \{\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_7, \ldots, \mathbf{x}_9\} \sim \{\mathbf{x}_3, \mathbf{x}_N, \mathbf{x}_2, \ldots, \mathbf{x}_1\} \sim \ldots$$

Thus for indistinguishable particles the configuration space is

$$\mathcal{C} = [(\mathbb{R}^D)^N - \Delta] \;/\sim$$

where "$/\sim$" means that we are "modding out" by the equivalence relationship $\sim$. This is just a fancy way to say that the order in which we list the coordinates $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_N\}$ does not matter (or as described in section 3.2, we choose some convention for the order, like always writing the left-most first). In the case of 2 identical particles above, this reduced the Hilbert space by a factor of 2. With N identical particles this will 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 parti-

[7] Mathematicians often write $(\mathbb{R}^D)^N \setminus \Delta$ to represent removing Δ from the set $(\mathbb{R}^D)^N$.

[8]The curves are directed because we do not allow them to double-back in time as shown in Fig. 3.2, that would represent particle-hole creation or annhilation, which we do not yet consider.

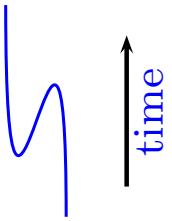


Fig. 3.2 A double-back in time is not allowed in our considerations here (and not allowed in the braid group) as it corresponds to creation and annihilation of particles at the turning around points.

[9]In fact what we really want is the *fundamental groupoid* which allows for the fact that the initial and final positions of particles may not be the same. However, for illustration, the fundamental group will be sufficient.

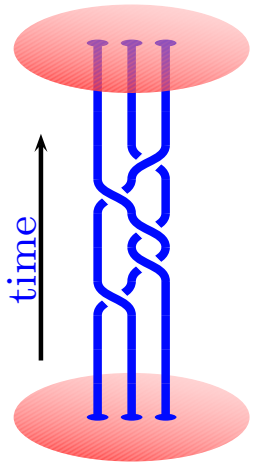


Fig. 3.3 A path through configuration space for 3 Particles in 2 dimensions (i.e, world lines in 2+1 D) is a braid with three strands.

[10]The identity element 1 of the braid group is everything that is topologically equivalent to the non-braid, i.e., particles that do not change their position in space at all. It is easy to see that $\sigma_i \sigma_i^{-1} = 1$.

cles 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[8]. Since we want to add up all of these possible paths in a path integral it is useful to try to understand the structure of this space of paths better.

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 assuming the endpoints are fixed. The group of paths through $\mathcal{C}$ is familiar to mathematicians and is known as the first homotopy group $\Pi_1(\mathcal{C})$ or fundamental group[9] (See section 33.3). 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.

3.3.1 Paths in 2+1 D, the Braid Group

A path through the configuration space of particles in 2 dimensions is known as a braid. An example of a braid is shown in Fig.3.3.

A few notes about braids:

(1) Fixing the endpoints, the braids can be deformed continuously, and so long as we do not cut one string through another, it still represents the same topological class, or the same element of the braid group.

(2) We cannot allow the strings to double-back in time as in Fig. 3.2. This would be pair creation or annihilation, which we will consider later, but not now.

The set of braids have mathematical group structure (See section 33.2): multiplication of two braids is defined by stacking the two braids on top of each other – first do one then do another. It is easy to see that braids can be decomposed into elementary pieces which involve either clockwise or counterclockwise exchange of one strand with its neighbor. These elementary pieces involving single exchanges are known as generators.

The braid group on N strands is typically notated as B_N. The generators of the braid group on 4 strands are shown in Fig. 3.4. Any braid can be written as a product of the braid generators and their inverses[10]. The "multiplication" of the generators is achieved simply by stacking the generators on top of each other. 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 listed right-most first), although sometimes people use the opposite convention! The important thing is to fix a convention and stick with it!

Note that many different braid words can represent the same braid. An example of this is shown for B_4 in Fig. 3.5. Although a braid can be written in many different ways, it is possible to define invariants of the braid which do not change under deformation of the braid — so long as

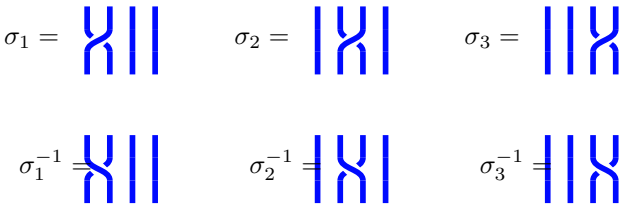


Fig. 3.4 The three generating elements $\sigma_1, \sigma_2, \sigma_3$ of the braid group on 4 strands, B_4, and their inverses $\sigma_1^{-1}, \sigma_2^{-1}, \sigma_3^{-1}$. Any braid on four strands (any element of B_4) can be written as a product of the braid generators and their inverses by simply stacking these generators together (See Fig. 3.5 for examples).

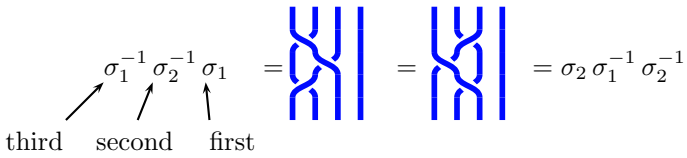


Fig. 3.5 Two braid words in B_4 that represent the same braid. The figure on the left can be continuously deformed to the one on the right, keeping endpoints fixed. The braidwords are read from right to left indicating stacking the generators from bottom to top.[11]

the braid is topologically unchanged. One very useful braid invariant is given by the so-called winding number

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

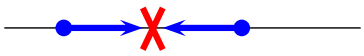
where an overcrossing is a σ and an undercrossing is a σ^{-1}. As can be checked in Fig. 3.5, the winding number is independent of the particular way we represent the braid. As long as we do not cut one strand through another or move the endpoints (or double-back strands) the winding number, a braid invariant, remains the same.

3.3.2 Paths in 3+1 D, the Permutation Group

We now turn to consider physics in 3+1 dimensions. A key fact is that it is not possible to knot a one-dimensional world-line that lives in a four-dimensional space. If this is not obvious consider the following lower dimensional analogue,[12] shown in Fig. 3.6. 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

[11] The observant reader will see the similarity here to Reidermeister moves of type-III discussed in section 2.6.1. Similarly $\sigma_i \sigma_i^{-1} = 1$ is a type-II move.

In one dimension:
Two objects cannot cross



In two dimensions:
Two objects can go around each other

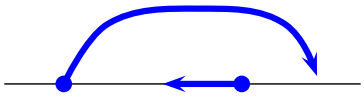


Fig. 3.6 **Top:** In one dimension, two points cannot cross through each other without hitting each other. **Bottom:** However, if we allow the points to move in two dimensions they can get around each other without touching. This is supposed to show you that one-dimensional world-lines cannot form knots in four-dimensional space.

[12] It would be very convenient to be able to draw a diagram in four dimensions!

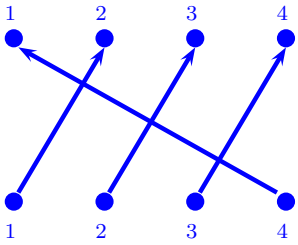


Fig. 3.7 Paths in 3+1 D are elements of the permutation group (or symmetric group) S_N (See section 33.2.1). Shown here is an element of S_4.

[13]One way to think about the relationship between the symmetric group and the braid group is to say that the symmetric group S_N is a "truncation" of the braid group B_N, meaning that it obeys the same group properties, except that in S_N, the element σ_i^2 has been identified with the identity.

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 discover that the strings can get by each other without ever touching each other! Hence there are no knots of one dimensional objects embedded in four dimensions.

Given that in 3+1 D 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 draw things that look a bit like braid-diagrams but now there is no meaning to an over or under-crossing. If the world line lives in 3+1 dimensions, everything can be unentangled without cutting any of the world lines until the diagram looks like Fig. 3.7: indicating only where lines start and end. This is precisely describing the permutation group, or symmetric group S_N (see section 33.2.1). Note that in the symmetric group an 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![13]

3.3.3 Building a Path Integral

We now return to the issue of building a path integral. We will follow the intuition we gained in the two particle case, but now we will include 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 construct the path integral as

$$\langle\{\mathbf{x}\}_f|\hat{U}(t_f,t_i)|\{\mathbf{x}\}_i\rangle = \mathcal{N}\sum_{g\,\in\,G}\rho(g)\sum_{\substack{\text{paths}\,\in\,g\\ i\to f}} e^{iS[\text{path}]/\hbar} \tag{3.5}$$

[14]In the nonabelian case discussed in section 3.5 below the ket $|\{\mathbf{x}\}\rangle$ is given an additional index to become $|n,\{\mathbf{x}\}\rangle$ with $n = 1\ldots M$. This then implies a basis choice for the M-dimensional space, and this basis choice for one set of positions $\{\mathbf{x}\}$ can be chosen independently of the basis choice for a different set of positions. When the initial and final positions are not the same we can make two independent basis choices and changing these choices simply pre- or post- multiplies the representation ρ by the appropriate basis changing unitaries. This caution is related to note 9 above.

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 is the braid group B_N for 2+1 dimensions. Here we have split the sum over paths into the different classes — the outer sum being a sum over the classes g and the inner sum being the sum over all paths of type g, i.e., a set of paths that can be continuously deformed into each other. We have also introduced[14] a factor of $\rho(g)$ out front where ρ is a *unitary representation* of the group G. (See section 33.2.4 on group theory).

To show that Eq. 3.5 is allowed by the laws of quantum mechanics, we need only check that it obeys the composition law – we should be able to construct all paths from i to f in terms of all paths from i to m and all paths from m to f.

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

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

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

So we have constructed all possible paths from i to f and split them into class g_2 in the region i to m and then class g_1 in the region m to f. When we compose these paths we will get a path of type $g_1 g_2$. The prefactors of the paths $\rho(g_1)$ and $\rho(g_2)$ then multiply and we get $\rho(g_1)\rho(g_2) = \rho(g_1 g_2)$ since ρ is a representation (the preservation of multiplication is the definition of being a representation! See section 33.2.4). So the prefactor of a given path from i to f is correctly given by $\rho(g)$ where g is the topological class of the path. In other words, the form shown in Eq. 3.5 is properly preserved under composition, which is what is required in quantum mechanics!

3.4 Abelian Examples

Let us consider the case where the representation ρ of our group G of paths through configuration space is one dimensional — in other words it is a mapping from g to a complex phase.[15]

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.

[15] We call these cases *abelian* since the group G is commutative.

3.4.1 3+1 Dimensions

In 3+1 D, the group G of paths through configuration space is the symmetric group S_N. It turns out that there are *only two possible*[16] one-dimensional representations of S_N:

- **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 considered above in section 3.2 where the exchange was assigned a -1.

[16] See exercise 3.2. This is a fairly short proof!

3.4.2 2+1 Dimensions

In 2+1 D, the group G of paths through configuration space 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 other words, 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 nonabelian 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 [1977][17], and popularized by Wilczek [1982][18]. Soon thereafter, Halperin [1984] and then Arovas, Schrieffer, and Wilczek [1984] showed theoretically that anyons really occur in fractional quantum Hall systems. We will examine these physical systems in detail starting in chapter **??**.

3.5 Nonabelian 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 correspond 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 numbers, like any "color" or "flavor" or "spin" 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 $n = 1 \ldots M$. An arbitrary wavefunction of the system can then be expressed as

$$|\psi_{\{\mathbf{x}\}}\rangle = \sum_{n=1}^{M} A_n |n;\{\mathbf{x}\}\rangle \tag{3.6}$$

with the A_n's being some complex coefficients.[19] 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

[17] There is no reason why this should not have been discovered in the 1930s, but no one bothered to think about it. It is a lucky coincidence that an experimental system of anyons was discovered so soon after the theoretical proposal (fractional quantum Hall effect, discovered by Tsui, Stormer, and Gossard [1982], see chapter ***), since the original theoretical work was entirely abstract, and they were not thinking about any particular experiment.

[18] Among other things, Wilczek coined the term *anyon*. (He also won a Nobel Prize for asymptotic freedom.)

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

M-dimensional representation of the braid group in our path integral! We thus identify that $\rho(g)$ in Eq. 3.5 is an M by M unitary matrix

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

which is a representation of G and must also be unitary so as to assure that probability is conserved. The propagator in Eq. 3.5 should now be thought of as a propagator between the initial ket $|n'; \{\mathbf{x}\}_i\rangle$ and the final bra $\langle n; \{\mathbf{x}\}_f|$. The unitary matrix $U(g)$ will act on the coefficients A_n (which is a vector) in Eq. 3.6.

Let us now consider the process shown in Fig. 3.8. Here an initial wavefunction is represented as shown in Eq. 3.6 as a vector $A_n^{(i)}$ multiplying basis states $|n; \{\mathbf{x}\}\rangle$ as in Eq. 3.6. We braid the particles around each other in some braid g and bring them back to the same positions. After braiding the wavefunction should still be composed of the same basis states $|n; \{\mathbf{x}\}\rangle$ since the particles are at the same positions and thus can be written in the form of Eq. 3.6 with a vector $A_n^{(f)}$. The final vector is obtained from the initial vector simply by multiplying by the unitary operator which is the representation of our braid group element g

$$A_n^{(f)} = [U(g)]_{n,n'} A_{n'}^{(i)} \tag{3.7}$$

A particle that obeys this type of braiding statistics is known as a **nonabelian anyon**, or **nonabelion**.[20] The word "nonabelian" means noncommutative, and the term is used since generically matrices (in this case the U matrices) don't commute.

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

$$M \sim \mathsf{d}^N \tag{3.8}$$

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) We will see in section 17.1 that (up to a possible sign) this quantum dimension d is actually the value d of the unknot[21].

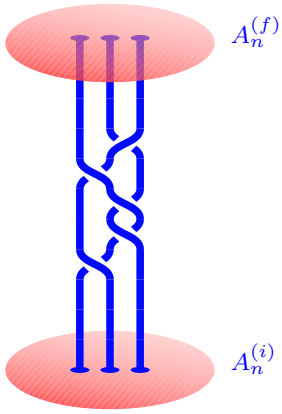


Fig. 3.8 An initial state is described by a vector $A_n^{(i)}$ multiplying the basis states $|n; \{\mathbf{x}\}_i\rangle$ as in Eq. 3.6. The particles are braided around each other in a braid g and brought back to the same positions. The final state is again described in terms of the same basis vectors but now with coefficients $A_n^{(f)}$ which are obtained from the initial vector by application of the unitary matrix $U(g)$ as shown in Eq. 3.7. Here $U(g)$ is a representation of the braid group.

[20] The idea of nonabelian anyons was explored first in the 1980s and early 90s by several authors in different contexts. Bais [1980] in the context of gauge theories; Fröhlich and Gabbiani [1990] and Fredenhagen et al. [1989] in very abstract sense; Witten [1989]; Chen et al. [1989] in the language of topological quantum field theories; and Moore and Read [1991] in the context of quantum Hall effect.

[21] Because of the possible sign, we distinguish the two quantities by using a different typeface.

[22] And initialization and measurement.

[23] 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.

Some Quick Comments on Quantum Computing:

Quantum Computing is nothing more than the controlled application of unitary operations to a Hilbert space[22]. Unitary operations is exactly 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.[23] In chapter 11 we will give some more explicit descriptions of how one does quantum computation by braiding anyons.

3.5.1 Parastatistics in 3+1 Dimensions

Is it possible to have exotic nonabelian statistics in 3+1 dimensions? Indeed, there do exist higher dimensional representations of the symmetric group, so one can think about particles that obey more complicated statistics even in 3+1 dimensions — which is often known as *parastatistics*. However, it turns out that, subject to some "additional constraints", it is essentially not possible to get anything fundamentally new — all we get is bosons and fermions and possibly some internal additional degrees of freedom. The proof of this statement is due to Doplicher et al. [1971, 1974] and took some 200 pages when it was first proven[24].

[24] A more concise derivation of the key portion of this result was given using modern category theory techniques by Müger [2007]. While this shorter proof is only 40 pages long, in order to understand the 40 pages you need to read a 400 page book on category theory first!

However, we should realize that in making statements like this, the fine print is important. As I mentioned in the previous paragraph we want to add some "additional constraints" and these are what really limit us to just bosons and fermions. What are these additional constraints?

(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 complicated structure that allows not just braiding particles around each other, but also creating and annhiliating and even merging particles by bringing them together. This structure is given by category theory, some parts of which we will encounter (in simplified language) starting in chapter 8.

(2) We also want some degree of locality. If we do an experiment on Earth, while off on Jupiter someone creates a particle-antiparticle pair, we would not want the particles on Jupiter 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 full details of how this happens. However, once we see the full structure of anyons in 2+1 dimensions, it ends up being fairly clear why 3+1 dimensions will be so restrictive. We return to this issue in section 20.3 where we will give further discussion.

We should note that despite this important result, 3+1 D 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). We would then need to consider the topology of the world-sheets describing loops moving through time.

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 nonabelian.
- Quantum computation can be performed by braiding with certain nonabelian representations.

Further Reading

- For more discussion of particle statistics, a nice albeit somewhat dated book is Wilczek [1990].
- A good review discussing many aspects of exotic statistics is Nayak et al. [2008].

For a basic primer on path integrals see

- R. MacKenzie, *Path Integral Methods and Applications*, https://arxiv.org/abs/quant-ph/0004090
- The classic reference on the subject is Feynman and Hibbs [1965].

Exercises

Exercise 3.1 About the Braid Group

(a) Convince yourself geometrically that the defining relations of the braid group on M particles B_M are:

$$\sigma_i\,\sigma_{i+1}\,\sigma_i = \sigma_{i+1}\,\sigma_i\,\sigma_{i+1} \qquad 1 \le i \le M-2 \tag{3.9}$$

$$\sigma_i\,\sigma_j = \sigma_j\,\sigma_i \quad \text{for} \quad |i-j|>1, \qquad 1 \le i,j \le M-1 \tag{3.10}$$

(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

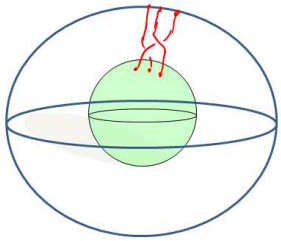


Fig. 3.9 An element of the braid group $B_3(S^2)$. The braid shown here is $\sigma_1\sigma_2^{-1}$

is useful to think of time as being the radial direction of the sphere, so that braids are drawn as in Fig. 3.9.

The braid generators on the sphere still obey Eqns. 3.9 and 3.10, but they also obey one additional identity

$$\sigma_1\sigma_2\ldots\sigma_{M-2}\sigma_{M-1}\sigma_{M-1}\sigma_{M-2}\ldots\sigma_2\sigma_1 = I \tag{3.11}$$

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. 3.11 is not quite enough. We will try to figure out below why this is so by using Ising Anyons as an example.]

Exercise 3.2 About the Symmetric Group

Show that Eqs. 3.9 and 3.10 also hold for the generators of the symmetric group S_M on M particles, where σ_i exchanges particle i and $i+1$. In the symmetric group we have the additional condition that $\sigma_i^2 = 1$. Prove the statement used in section 3.4.1 that there are only two one-dimensional representations of the symmetric group. Hint: The proof is just a few lines. Use $\rho(\sigma_i)\rho(\sigma_j) = \rho(\sigma_i\sigma_j)$ where ρ is a representation.

Exercise 3.3 Ising Anyons and Majorana Fermions

The most commonly discussed type of nonabelian 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 nonabelian 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

$$\{\gamma_i, \gamma_j\} \equiv \gamma_i\gamma_j + \gamma_j\gamma_i = 2\delta_{ij} \tag{3.12}$$

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 *** (Ising F matrix)

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

$$U_{ij} = \frac{e^{i\alpha}}{\sqrt{2}}\left[1 + \gamma_i\gamma_j\right] \qquad i < j. \tag{3.13}$$

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. 3.9 and 3.10 yields equalities. This representation is 2^N dimensional since the ground state degeneracy is 2^N.

(c) Consider the operator

$$\gamma^{\text{FIVE}} = (i)^N \gamma_1 \gamma_2 \dots \gamma_{2N} \tag{3.14}$$

(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. 3.11, one must choose the right abelian phase α in Eq. 3.13. 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. 3.11 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. 3.11. 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 3.4 Small Numbers of Anyons on a Sphere

On the plane, the braid group of two particles is an infinite group (the group of integers describing the number of twists!). 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 Effect and Charge-Flux Composites

4

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 nonabelian generalization.

4.1 Review of Aharanov-Bohm Effect

Let us consider the two slit interference experiment shown in Fig. 4.1. We all know the result of the two slit experiment but let us rewrite the calculation in the language of 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 on the output screen, and k is the wavevector of the incoming wave. In other words, we get the usual two slit calculation pioneered by Thomas Young in the early 1800s.

Now let us change the experiment to that shown in Fig. 4.2. Here we assume the particle being sent into the interferometer is a charged particle, such as an electron. In this case a magnetic field is added inside the middle box between the two paths. No magnetic field is allowed to leak out of the box, so the particle never experiences the 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 [1949], then re-predicted independently by Aharonov and Bohm [1959][1].

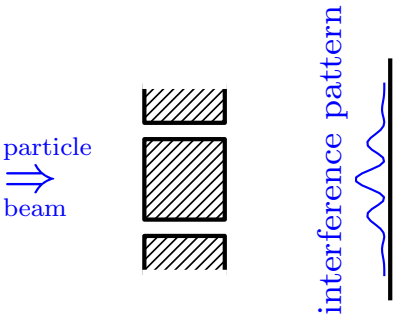


Fig. 4.1 The Young two slit experiment (not to scale).

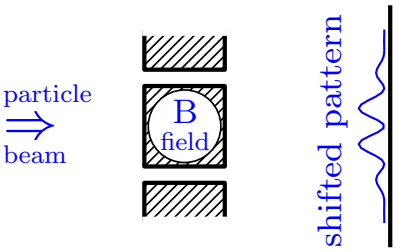


Fig. 4.2 Adding a magnetic field inside the middle box in the Young two slit experiment. Here the circular region includes a constant magnetic field. No magnetic field leaks out of the box. Nonetheless, if the particle being sent into the interferometer is charged, the interference pattern is changed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

[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 as much how strange it is! The first experimental observation of the effect was by Chambers [1960], although many more careful experiments have been done since.

So why does this strange effect occur? There are several ways to understand it, but for our purpose it will be best 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

$$\begin{aligned} \mathbf{B} &= \nabla \times \mathbf{A} \\ \mathbf{E} &= -\nabla A_0 - d\mathbf{A}/dt \end{aligned}$$

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

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

where q is the particle charge. It is an easy exercise to check that the Euler-Lagrange equations of motion that result from this Lagrangian correctly gives motion under the Lorentz force as we should expect for a charged particle in an electromagnetic field.[2]

We are interested in a situation where we add a static magnetic field to the system. Thus, we need only include $q\mathbf{A}(\mathbf{x}) \cdot \dot{\mathbf{x}}$ in the Lagrangian. The action then gets changed by

$$S \to S_0 + q \int dt\ \dot{\mathbf{x}} \cdot \mathbf{A} = S_0 + q \int \mathbf{dl} \cdot \mathbf{A} \tag{4.2}$$

where S_0 is the action in the absence of the magnetic field and the integral on the far right is a line integral along the path taken by the particle.

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 again 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 difference is given by

$$\exp\left[\frac{iq}{\hbar}\int_{\text{slit 1}} \mathbf{dl}\cdot\mathbf{A} - \frac{iq}{\hbar}\int_{\text{slit 2}} \mathbf{dl}\cdot\mathbf{A}\right] = \exp\left[\frac{iq}{\hbar}\oint \mathbf{dl}\cdot\mathbf{A}\right] \tag{4.3}$$

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

Using Stokes' theorem, we have

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

[2] Here are the steps: Start with the Euler-Lagrange equations

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}_k} = \frac{\partial L}{\partial x_k}.$$

This gives us

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

So that

$$m\ddot{x}_k = q(\mathbf{E} + \dot{\mathbf{x}}\times\mathbf{B})_k \quad .$$

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 original Lagrangian Eq. 4.1 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 multiple 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.

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

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

4.2 Anyons as Charge-Flux 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. 4.3. We will notate this charge-flux composite object as a (q, Φ) particle. If we drag one such particle around another, we then accumulate a phase due to the Aharanov-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

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

Thus we have (as shown in Fig. 4.4)

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

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 abelian anyons as discussed in chapter 3.

Note that the same type of calculation would show us that taking a composite particle with charge q_1 and flux Φ_1 all the way around a composite particle with charge q_2 and flux Φ_2 would accumulate a phase of $e^{i\varphi}$ with $\varphi = (q_1\Phi_2 + q_2\Phi_1)/\hbar$.

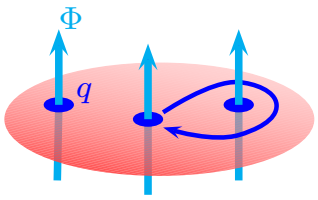


Fig. 4.3 Abelian anyons represented as charges bound to flux tubes through the plane. The charge of each particle is q, the flux of each tube is Φ. Dragging one particle around another incurs a phase both because charge is moving around a flux, but also because flux is moving around a charge.

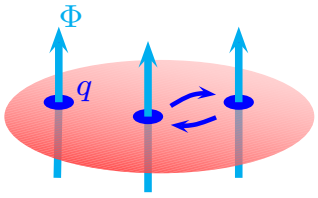


Fig. 4.4 An exchange. Two exchanges is the same as dragging one particle all the way around the other as shown in Fig. 4.3.

[3]Almost any prescription for attaching flux to charge (for example, break the flux into four pieces and attach one piece on each of four side of the charge) will give the same result. However, if we try to put the flux and charge at exactly the same position, we get infinities that we don't know how to handle!

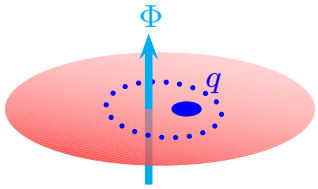


Fig. 4.5 Tying flux to charge. We put the flux and the charge at slightly different positions. As a result, when we rotate the particle around its own axis a phase is accumulated as the charge and flux go around each other.

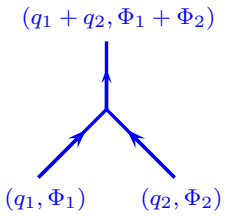


Fig. 4.6 Fusing two anyons to get an anyon of a different type which has the sum of fluxes and the sum of charges.

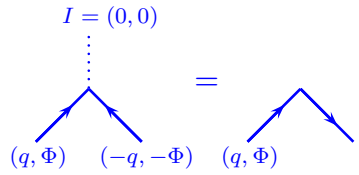


Fig. 4.7 Fusing an anyon and an antianyon to get the vacuum (I) drawn as dotted line. Note that the antianyon moving forward in time is drawn as a downpointing arrow — which looks like an anyon moving backwards in time.

[4]The vacuum or identity particle can be denoted e, or I or 0 or 1 depending on the context. This nomenclatural problem stems from a similar problem in group theory, see section 33.2.

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 the charge and flux at very slightly different positions as shown in Fig. 4.5[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 of Fig. 2.7).

4.2.1 Fusion of Anyons

We can consider pushing two anyons together to try to form a new particle. We expect that the fluxes will add and the charges will add. This makes some sense as the total charge and total flux in a region should be conserved (this is an important principle that we will encounter frequently!). We sometimes will draw a "fusion diagram" as in Fig. 4.6 to show that two anyons have come together to form a composite particle.

A simple example of this is pushing together two particles both having the same charge and flux (q, Φ). In this case 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 $\theta = 4q\Phi/\hbar$ (since the factor of 2 in charge multiplies the factor of 2 in flux!).

4.2.2 Anti-Anyons and the Vacuum Particle

We now introduce the concept of an anti-anyon. This is a charge-flux composite which instead of having charge and flux (q, Φ) has charge and flux $(-q, -\Phi)$. Fusing an anyon with its anti-anyon results in pair annihilation — the two particles come together to form the vacuum (which we sometimes[4] refer to as the identity I) which has zero total charge and zero total flux, as shown in Fig. 4.7. It may seem a bit odd to call the absence of any charge or any flux a "particle". However, this is often convenient since it allows us to think of pair annihilation (as in the left of Fig. 4.7) in the language of fusion.

In the right of Fig. 4.7 we show that it is sometimes convenient *not* to indicate the vacuum particle. In this case, we have written the anti-anyon moving forward in time as an anyon moving backwards in time.

If the phase of dragging an anyon clockwise around an anyon is 2θ, then the phase of dragging an anti-anyon clockwise around an anti-anyon is also 2θ. (The two minus signs on the two anyons cancel — negative

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

4.3 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[5].

We can draw the torus as a square with opposite edges identified as shown in Fig. 4.8. 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-antiparticle 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-antiparticle 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.

[5] See note 1 in chapter 33.

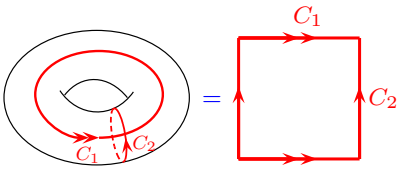


Fig. 4.8 Drawing a torus as a rectangle with opposite edges identified. The two noncontractable cycles around the torus can be considered to be the edges of the square, labeled C_1 and C_2 here.

Both of these operators are unitary because they can be implemented (in principle) with some time-dependent Hamiltonian[6]. 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 procedure is shown in Fig. 4.9.

So what we have now is two operators T_1 and T_2 which do not commute with each other. Indeed, we have[7]

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

But both T_1 and T_2 commute with the Hamiltonian (since they start and end with states of exactly the same energy[8]). 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

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

Note that we are labeling the ket $|\alpha\rangle$ by its eigenvalue under the ap-

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

[7] 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.

[8] Strictly speaking this means they commute with the Hamiltonian within the ground state space, or equivalently the commutators $[T_1, H]$ and $[T_2, H]$ both annihilate the ground state space.

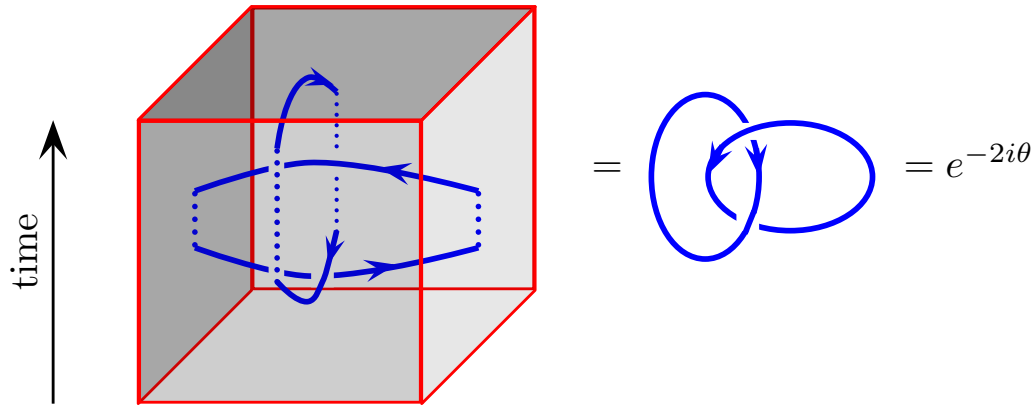


Fig. 4.9 The torus is drawn as a horizontal rectangle with opposite ends identified. Time runs vertically. First create a particle-antiparticle pair at the center of the rectangle and move them in opposite directions, right and left, until they meet at the edges of the rectangle to reannhiliate. Note that a particle moving to the right or an antiparticle moving to the left are both drawn as a rightpointed arrow. Next create a particle-antiparticle pair in the center of the torus and move them 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 to make the figure on the left look like the linked rings, we should not quite annihilate the particles at the end of the first and second step (turning the dotted lines into solid lines). This is allowed since bringing a particle-anti-particle pair close together looks like they have fused together to the vacuum if we view it from far away.

plication of T_1. 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 since T_2 commutes with the Hamiltonian. But now

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

This new ground state $T_2|\alpha\rangle$ has eigenvalue $e^{i\alpha+2i\theta}$ under application of T_1. We thus 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). Starting with the ground state $|\alpha\rangle$ we can generate a series of ground states by successive application of T_2,

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

When we try to generate yet another state, we get the phase $\alpha+2\pi$ which is equivalent to α since it is describing a complex phase, so we are back to the original state. So we now have m independent ground states.[9] Note in particular that the ground state degeneracy *of the system with*

[9]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 integer, where one generally expects that integer to be 1 but there could be additional accidental degeneracy.

no anyons in it is related to the statistical angle θ of the anyons if they were to be created.

4.3.1 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 of the multiple ground states

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

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 unlikely when the energy gap for creating excitations is large[10]. Hence the quantum superposition is "topologically protected".

In fact, one does not even need to have a system on a torus in order to have a degenerate ground state. It is often sufficient to have an annulus geometry (a disk with a big hole in the middle as shown in Fig. 4.10). In this case, T_1 could correspond to moving an anyon around the loop of the annulus and T_2 could correspond to moving an anyon from the inside to the outside edge.[11]

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 (See exercise 4.1). Thus by using high genus one can obtain very very large Hilbert spaces in which to store quantum information.

[10] Strictly speaking, at any finite temperature for any size system there is a finite time for this process to occur, although it might be very long.

[11] In this case it is often not precisely true that the ground states are entirely degenerate (since there is a non-zero 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. A classic example of this is discussed by Gefen and Thouless [1993].

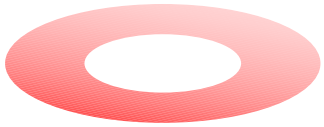


Fig. 4.10 An annulus.

4.3.2 Number of Species of Anyons

Having established multiple vacuum states on a torus, let us now return to study the anyons that we could create in such a system. Again let us consider anyons of statistical angle $\theta = \pi p/m$ with p and m relatively prime. We can describe such anyons[12] with a charge-flux composite $(q, \Phi) = (\pi p/m, 1)$. Fusion of n of these elementary anyons will have charge and flux given by[13]

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

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

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

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

[13] It is only a slight abuse of notation to write the ket $|\text{“}n\text{”}\rangle$ to mean a cluster of n elementary anyons.

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[14] 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 $|\text{“}m\text{”}\rangle$. Sometimes one says that $|\text{“}m\text{”}\rangle$ is a *transparent* particle.

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, as in Fig. 4.4) a $|\text{“}m\text{”}\rangle = (\pi p, m)$ particle with another $|\text{“}m\text{”}\rangle = (\pi p, m)$ particle. To see this note that exchange gives a phase πpm since it is half of the $2\pi pm$ phase for wrapping one particle all the way around the other (as in Fig. 4.3). 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 is a bit more difficult to handle[15].

Neglecting this more complicated case with transparent particles, we are correct to conclude that we have exactly m different species of anyons – and also m different ground states on the torus as calculated above. 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.

[14] As mentioned at the beginning of section 4.2 the total phase is given by $q_1\Phi_2 + q_2\Phi_1 = (n\pi p/m)m + (\pi p)n$.

[15] Whenever we have a particle that braids trivially with all other particles (i.e., is transparent), the theory is more complicated. Later on we will call this kind of theory “non-modular.” See section 17.3.1.

Chapter Summary

- The Charge-Flux composite model describes abelian anyons — with the braiding phase coming from Aharonov-Bohm effect.
- We introduced idea of fusion, antiparticles, and spin
- The vacuum for a system of anyons is nontrivial and can be a quantum memory.

Further Reading

A good reference for the charge-flux composite model is John Preskill's lecture notes (Preskill [2004]).

Exercises

Exercise 4.1 Abelian Anyon Vacuum on a Two-Handle Torus

Using similar technique as in section 4.3, show that the ground state vacuum degeneracy on a two handle torus is m^2 for a system of abelian anyons with statistical angle $\theta = \pi p/m$ for integers p and m relatively prime. Hint: Consider what the independent cycles are on a two-handled torus and determine the commutation relations for operators T_i that take anyon-antianyon pairs around these cycles.

Chern-Simons Theory Basics

5

5.1 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, so-called "Chern-Simons" field theory[1], is the main 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 imagining a gauge field a_α, known as the Chern-Simons vector potential, analogous to the vector potential A_α we know from regular electromagnetism. 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

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

Here we have written L_0 to be the Lagrangian of our particles without considering the coupling to the (Chern-Simons) vector potential. This might 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

$$\mathcal{L} = \frac{\mu}{2}\epsilon^{\alpha\beta\gamma}a_\alpha\partial_\beta a_\gamma - j^\alpha a_\alpha \tag{5.1}$$

where j^α is the particle current, μ is some coupling constant, and ϵ is the antisymmetric tensor[2]. The indices α, β, γ take 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. 5.1 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. 5.1 couples the Chern-Simons vector potential to the particles in the system. Its form, $j^\alpha a_\alpha$, may look unfamiliar but it is actually just the expected coupling of the charged

[1] 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 20 billion dollars (as of 2018). 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.

[2] The antisymmetric tensor is given by $\epsilon^{012} = \epsilon^{120} = \epsilon^{201} = 1$ and $\epsilon^{210} = \epsilon^{102} = \epsilon^{021} = -1$.

particles to a vector potential analogous to what we used when we discussed Aharonov-Bohm effect in section 4.1. To see this, let us carefully define the particle current j^α. If we have N particles then the current is

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

[3]Again not the real electromagnetic charge, but rather the charge that couples to the Chern-Simons vector potential a_α. Later in this chapter we will set $q=1$ along with $\hbar=1$ for simplicity of notation.

The j^0 component, the charge density[3], is just a delta function peak at the position of each particle with value given by the particle charge q. The 1 and 2 component, $\mathbf{j}$ is a delta function at the position of each particle with prefactor given by the velocity of the particle times its charge. Now when $-j^\alpha a_\alpha$ is integrated over all of space we get

$$\sum_{n=1}^{N} q_n \left[\mathbf{a}(\mathbf{x}_n)\cdot\dot{\mathbf{x}}_n - a_0(\mathbf{x}_n)\right] \tag{5.2}$$

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

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 (one measures only the flux enclosed, as with the Aharonov-Bohm effect). The Chern-Simons term (the first term in Eq. 5.1) is also gauge invariant, at least on a closed manifold if we can integrate by parts. To see this, make an arbitrary gauge transformation

$$a_\mu \to a_\mu + \partial_\mu \chi \tag{5.3}$$

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 antisymmetry. Note that 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 term does we need to look at the Euler-Lagrange equations of motion. We have

$$\frac{\partial\mathcal{L}}{\partial a_\alpha} = \partial_\beta\left(\frac{\partial\mathcal{L}}{\partial(\partial_\beta a_\alpha)}\right) \tag{5.4}$$

[4]It may look like the right result would have $\mu/2$ on the right hand side, given that it is $\mu/2$ in Eq. 5.1. However, note that when we differentiate with respect to a_α on the left hand side of Eq. 5.4, we also generate an identical factor of $\mu/2$ and these two add up.

which generates the equations of motion[4]

$$j^\alpha = \mu\epsilon^{\alpha\beta\gamma}\partial_\beta a_\gamma \tag{5.5}$$

This equation of motion demonstrates flux binding. To see this, let us

look at the 0th component of this equation. We have

$$j^0 = \sum_{n=1}^{N} q_n \delta(\mathbf{x} - \mathbf{x}_n) = \mu(\nabla \times \mathbf{a}) = \mu b \tag{5.6}$$

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

For simplicity, let us now assume all particles are identical with the same charge $q_n = q$. We might expect that the phase obtained by exchanging two such identical charges would be given by the charge times the flux or $\theta = q^2/\mu$ analogous to section 4.2. 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. 5.5 by a_α and then plug it back into[5] the Lagrangian 5.1. We then end up with

$$\mathcal{L} = -\frac{1}{2} j^\alpha a_\alpha$$

In other words, the Lagrangian 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

$$\sum_{\text{paths } \{\mathbf{x}(t)\}} \sum_{\text{all } a_\mu(\mathbf{x},t)} e^{i(S_0 + S_{CS} + S_{coupling})/\hbar} \tag{5.7}$$

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 that this means we should sum over 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

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

Formally when we write a functional integral we mean[6] 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.

At least formally we can thus rewrite Eq. 5.7 as

$$\sum_{\text{paths } \{\mathbf{x}(t)\}} e^{iS_0/\hbar} \int \mathcal{D} a_\mu(x)\ e^{iS_{CS}/\hbar}\ e^{i(q/\hbar)\int_{paths} dl^\alpha a_\alpha} \tag{5.8}$$

[5] One might worry about whether we are actually allowed to plug the equations of motion back into the Lagrangian when we do a full path integral, as in Eq. 5.7, where we are supposed to integrate over all field configurations, not just those that satisfy equations of motion. While generally in field theory one should not plug equations of motion back into the Lagrangian, it is actually allowed in this case because the Lagrangian is linear in each a_μ. For example, classically we can think of a_0 as being a Lagrange multiplier which enforces Eq. 5.6. Similarly in the functional integral when we integrate out a_0 it enforces that equation of motion as a strict constraint.

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

where S_0 is the action of the particles following the path but not interacting with the gauge field, S_{CS} is the action of the Chern-Simons gauge field alone (from the first term in Eq. 5.1). The final exponential in Eq. 5.8 represents the coupling (from the second term of Eq. 5.1) 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. 5.2). 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 5.2.

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

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

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!

Vacuum Abelian Chern-Simons Theory

Something we have pointed out above in section 4.3 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[7].

If we are consider a three dimensional manifold of the form $\mathcal{M} = \Sigma \times S^1$ for a 2D manifold Σ and S^1 represents time (compactified[8]) this integral gives exactly the ground state degeneracy of the system. 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 5.3.2). 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.

[7] 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 chapter 6 and Fig. 6.1.

[8] 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.

5.2 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 abelian Chern-Simons theory by promoting the gauge field a_α to be not just a vector of numbers, but rather a vector of matrices.[9] More precisely, to construct a nonabelian Chern-Simons theory, we consider a vector potential that takes values in a Lie algebra[10]. For example, if we choose to work with the Lie algebra of $SU(2)$ in the fundamental representation we can write a general element of this algebra as a sum of the three generators (proportional to $\sigma_x, \sigma_y, \sigma_z$) so that our Lie algebra valued gauge field is then[11]

$$a_\mu(x) = a^a_\mu(x) \left(\frac{\sigma_a}{2i}\right) \tag{5.9}$$

where σ_a are the Pauli matrices. 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 need to think about is the Wilson loop operators[12]

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

where here the integral follows some closed path L. This object, being the exponential of an integral of a vector potential, is essentially the nonabelian analogue[13] of the Aharonov-Bohm phase of Eq. 4.3). In Eq. 5.10, the P symbol indicates path ordering — analogous to the usual time ordering of quantum mechanics. The complication 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

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

where $x_1, x_2, x_3, \ldots$ are the small steps along the path.

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

$$a_\mu \to U^{-1} a_\mu U + U^{-1}\partial_\mu U \tag{5.12}$$

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[14] of the gauge transformation in Eq. 5.3. To see that this gauge transformation leaves the Wilson loop operators invariant (and hence is the right way to define a gauge transformation!) see section 5.3.1.

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

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

[9] If you have studied Yang-Mills theory, you already know about nonabelian vector potentials.

[10] See the introduction to Lie groups and Lie algebras in section 33.2.3. In brief: A Lie group is a group which is also a continuous manifold. A Lie *algebra* is the algebra of infinitesimal changes in this group. A prime example is the Lie group $SU(2)$ with algebra generated by $i\sigma_j$ with σ_j's being the Pauli operators. We write group elements as exponentials of the algebra $g = e^{i\boldsymbol{\sigma}\cdot\mathbf{n}}$.

[11] For general Lie algebras, we want to write $a_\mu = a^a_\mu T_a$ where T_a are the anti-Hermitian generators of the Lie algebra with $T_a = -T_a^\dagger$. This means that $[T_a, T_b] = f^{abc}T_c$ with f the so-called structure constants of the Lie group, and $\mathrm{Tr}[T_aT_b] \equiv -\frac{1}{2}\delta_{ab}$. In case of $SU(2)$ in the fundamental representation we have $T_a = -i\sigma_a/2$ with $f^{abc} = \epsilon^{abc}$. Be warned that other normalization conventions do exist, and changing conventions will insert seemingly random factors of 2 or i or worse.

[12] 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 a 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.

[13] The factor of i we usually have in the exponential of the Aharonov-Bohm phase (Eq. 4.3) is missing because it has been absorbed into a_μ in Eq. 5.9 (See comment in note 11). The factors of q and $\hbar$ are missing because we have set them to one as every theorist should do.

[14] Here take $U = e^{i\chi}$ and note that a factor of i is absorbed into the vector potential as mentioned in note 13.

Note that the second term in the brackets would be zero if the a_α were commutative. (In the abelian case above, we have no such term! See Eq. 5.1).

The Chern-Simons action is metric independent, which we show explicitly in the chapter appendix section 5.3.2. This means that space and time can be deformed continuously and the value of the action does not change. While this may not be obvious from looking at the form of the action, a large hint is that the action is written without any reference to the usual space-time metric $g_{\mu\nu}$.

Since Chern-Simons theory is also a gauge theory, we would like the action to be gauge invariant. It turns out that the action is *almost* gauge invariant, as we will discuss momentarily. At any rate it is close enough to gauge invariant to be of use for us!

It turns out that the Chern-Simons action is actually unique in being both metric independent and also (at least almost) gauge invariant. In 2+1 dimensions, no other action can be written down which involves only one gauge field and has these two properties: topological invariance and gauge invariance. This is what makes Chern-Simons theory such a crucial paradigm for topological theories in 2+1 dimensions.

Let us now return to this issue of how the Chern-Simons action is only *almost* gauge invariant. First of all, if the manifold has a boundary, we will run into non-gauge invariant terms as mentioned below Eq. 5.3. For now, let us just assume that our manifold has no boundaries.

More crucially there is another issue with gauge invariance. Under gauge transformation (at least on a closed manifold) as in Eq. 5.12 the Chern-Simons action transforms to (See exercise 5.2)

$$S_{CS} \to S_{CS} + 2\pi\nu k \tag{5.14}$$

where

$$\nu = \frac{1}{24\pi^2}\int_{\mathcal{M}} d^3x\ \epsilon^{\alpha\beta\gamma}\ \mathrm{Tr}\left[(U^{-1}\partial_\alpha U)(U^{-1}\partial_\beta U)(U^{-1}\partial_\gamma U)\right] \tag{5.15}$$

Surprisingly the complicated expression in Eq. 5.15 (sometimes known as the Pontryagin index) is always an integer (See section 5.3.3 for more detail). The integer ν gives the winding number of the map $U(x)$ from the manifold into the gauge group[15].

It may now look problematic that our Chern-Simons action is not a true gauge invariant (Eq. 5.14), but we note that the only thing entering our functional integral is $e^{iS_{CS}}$, not the Chern-Simons action itself. Thus, so long as we choose k, the so-called "level", as an integer (and since the winding number ν is also an integer), then we have a well defined functional integral of the form

$$Z(\mathcal{M}) = \int_{\mathcal{M}} \mathcal{D}a_\mu(x)\ e^{iS_{CS}} \tag{5.16}$$

where the result $Z(\mathcal{M})$ turns out to be a manifold invariant (see chapter appendix, section 5.3.2), meaning that smooth deformations of space and

[15] In the case of the gauge group being $SU(2)$, as mentioned in section 33.2.3, 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 $x \in S^3$ (space) to $U(x) \in S^3$ (group). 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. The case of zero winding number is anything that can be continuously deformed to $U = 1$ everywhere. However, we also can consider the identity mapping that S^3 (space) maps into S^3 (group) in the obvious way (every point goes to itself) 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. (See exercise 5.3)

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. The fact that the result should be a topological invariant should not be surprising given the fact that the Chern-Simons action itself is metric independent and therefore independent under deformations of space and time[16]. 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 33.1.1 for definition of S^3).

So for example, to find the link invariant corresponding to the two linked strings in Fig. 5.1, we have

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

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 (working with the spin 1/2 representation of the group, i.e, with Pauli matrices) we obtain the Kauffman invariant of the knot with $A = -(-i)^{(k+1)/(k+2)}$.

If we keep the same gauge group, but work with a different representation (for example, spin 1, rather than spin 1/2 in Eq. 5.9), we will obtain different "particle types" of the theory.

One can also choose to work with different gauge groups. Using $SU(N)$ and choosing a level k one obtains the two parameter HOMFLY knot polynomial (the two parameters here being N and k). Similarly, using $SO(N)$ at level k gives a two parameter Kauffman polynomial (not to be confused with the Kauffman bracket). Typically a Chern-Simons theory with gauge group G at level k is notated as G_k (For example, using $SU(2)$ at level 2 we write the theory as $SU(2)_2$). Changing the sign of k corresponds to taking the "mirror image" of the theory (the partition function is complex conjugated).

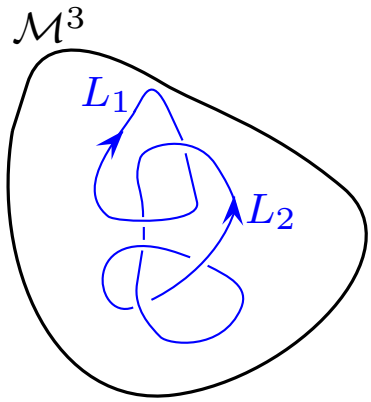


Fig. 5.1 A cartoon of a 3 manifold with a link made of two strands embedded in it.

[16] 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 2.6.1.) In field theory language this enters the calculation by how a point-splitting regularization is implemented.

5.3 Appendix: Odds and Ends about Chern Simons Theory

5.3.1 Gauge Transforms with Nonabelian Gauge Fields

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

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

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

$$W_C(x,y) \to U(x)^{-1}\ W_C(x,y)\ U(y) \tag{5.17}$$

Clearly this obeys composition of paths, and will correctly give a gauge invariant result for a closed Wilson loop. Now let us see what is required for the gauge field a_μ such that Eq. 5.17 holds. We consider

$$W_C(x, x+dx) = 1 + a_\mu dx^\mu \tag{5.18}$$

and its transformation should be

$$\begin{aligned} W_C(x, x+dx) &\rightarrow U(x)^{-1} W_C(x, x+dx) U(x+dx) \\ &= U(x)^{-1}[1 + a_\mu dx^\mu] U(x+dx) \\ &= U(x)^{-1}[1 + a_\mu dx^\mu][U(x) + dx^\mu \partial_\mu U(x)] \\ &= 1 + [U^{-1} a_\mu U + U^{-1}\partial_\mu U] dx^\mu \end{aligned} \tag{5.19}$$

By comparing Eq. 5.18 and Eq. 5.19 we see that the gauge transform rule Eq. 5.12 correctly gives a gauge invariant Wilson loop operator.

5.3.2 Chern Simons Action is Metric Independent

You will often see books state that Eq. 5.13 must be metric independent because you don't see the metric $g_{\mu\nu}$ written anywhere. 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 being the integral of a 3-form.

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. In other words, we are allowed to bend and stretch the space-time manifold, but the flux through a loop should stay constant. Under reparametrization of coordinates we have

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

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

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

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 three-dimensional Chern-Simons action 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!

5.3.3 Winding Number: The Pontryagin Index

We would like to show that the integral in Eq. 5.15 is indeed always an integer. While doing this rigorously is difficult, it is not too hard to see roughly how it must be done. First, we note that, like the Chern-Simons action, it is the integral of a three form so it does not care about the metric on the manifold (this is not surprising being that this winding number arose from the Chern-Simons action). One can then reparameterize the manifold in terms of coordinates within the group, and convert the integral over space into an integral over the group. The only thing that is left unclear is then in the mapping $U(x) : \mathcal{M} \to G$ how many times the group is covered in this mapping. We then have immediately that the given definition of the winding number must be an integer times some constant. By construction of a few examples, one can see that the constant is indeed unity (See exercise 5.4). A more detailed discussion of this issue is given in Vandoren and van Nieuwenhuizen [2008] and Rajaraman [1982].

5.3.4 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[17]). 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 orientable 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!)

[17] A detailed discussion of 2-framing is given by Atiyah [1990b]; Kirby and Melvin [1999]. This is fairly mathematical stuff!

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 time reversal invariant 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-dimensional boundaries and the other chirality ends up on the other D dimensional boundary[18]. 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 obviously won't care about the phase anyway!

[18] 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. See chapter ***.

5.3.5 Chern Simons Theory as Boundary of a Four Dimensional Topological Theory

With the considerations of the previous section 5.3.4, it is interesting to express Chern-Simons theory as the boundary theory of a 4D topological theory. To do this let us define the field strength tensor

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu]$$

This definition matches the expression for the electomagnetic field strength in the case where the fields are abelian such that the commutator vanishes. However, more generally for nonabelian gauge theories (including Yang-Mills theory) the additional commutator term must be added.

In 4D we can define the dual field strength

$${}^*F^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\lambda\rho}F_{\lambda\rho}$$

where ϵ is the antisymmetric tensor. We now consider the following topological action on a 4D manifold $\mathcal{N}$

$$S = \frac{\theta}{16\pi^2}\int_{\mathcal{N}} d^4x \ \mathrm{Tr}\,[F_{\mu\nu}{}^*F^{\mu\nu}]$$

This 4D action is well defined and non-anomalous, meaning it can be regularized and/or treated properly on a lattice.

With a bit of algebra the action can be rewritten as

$$S = \frac{\theta}{8\pi^2}\int_{\mathcal{N}} d^4x\ \partial_\mu G_\mu$$

where

$$G_\mu = 2\epsilon_{\mu\nu\lambda\rho}\,\mathrm{Tr}\left[A_\nu\partial_\lambda A_\rho + \frac{2}{3}A_\nu A_\lambda A_\rho\right] \tag{5.20}$$

Since the action can be written as the integral of a total derivative, it should give zero when integrated over a closed manifold $\mathcal{N}$. However when the manifold has a boundary one obtains

$$S = \frac{\theta}{8\pi^2}\int_{\partial N} d^3x\ G_\mu v^\mu$$

where v^μ is the unit vector normal to the boundary. Examining the form of Eq. 5.20 we realize that the action is precisely the Chern-Simons action on the 3D boundary manifold ∂N.

5.3.6 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. Working in the gauge $a_0 = 0$, in the Chern-Simons Lagrangian terms like $\partial_0 a_y$ multiply a_x and vice versa[19]. 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 different handles of a torus

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

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 identity that

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

which holds when $[A, B]$ is a number not an operator. This then 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 4.3 which created

[19] Note that for nonabelian Chern-Simons theories working in the $a_0 = 0$ gauge makes the a^3 term of the action vanish!

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 model can be realized in an abelian Chern-Simons theory.
- We introduced some ideas of general nonabelian Chern-Simons theory, including manifold invariants and turning Wilson loop operators into knot invariants.

A good reference for abelian Chern-Simons theory is

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

Some good references on nonabelian Chern-Simons theory are

- 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.
- *Current Algebras and Anomolies*, by S. Treiman, R. Jackiw, B. Zumino, and E. Witten (World Scientific) 1985. See particularly the chapters by R. Jackiw.
- G. Dunne, *Aspects of Chern-Simons Theory* in Topological aspects of low dimensional systems. Les Houches - Ecole d'Ete de Physique Theorique, vol 69. Springer, Berlin, Heidelberg, eds A. Cometet, T, Jolicoeur and S. Ouvry. Also available as arXiv:hep-th/9902115.

Exercises

Exercise 5.1 Polyakov Representation of the Linking Number

Consider a link made of two strands, L_1 and L_2. Consider the double line integral

$$\Phi(L_1, L_2) = \frac{\epsilon_{ijk}}{4\pi} \oint_{L_1} dx^i \oint dx^j \frac{x^k - y^k}{|\mathbf{x} - \mathbf{y}|^3}$$

(a) Show that Φ is equal to the phase accumulated by letting a unit of flux run along one strand, and moving a unit charged particle along the path of the other strand.

(b) Show that the resulting phase is the topological invariant known as the linking number — the number of times one strand wraps around the other, see section 2.6.2.

This integral representation of linking was known to Gauss.

Exercise 5.2 Gauge Transforming the Chern-Simons Action

Make the gauge transform Eq. 5.12 on the Chern-Simons action 5.8 and show that it results in the change 5.14. Note that there will be an additional term that shows up which it a total derivative and will therefore vanish when integrated over the whole manifold $\mathcal{M}$.

Exercise 5.3 Winding Numbers of Groups in Manifolds

Consider the mapping of $U(x) \in SU(2) \to S^3$. Construct an example of a map with winding number n for arbitrary n. I.e., find a representative of each group element of $\Pi_3(SU(2))$ (See note 15).

Exercise 5.4 Quantization of Winding Number

Let us consider the manifold S^3 which we consider as $\mathbb{R}^3$ plus a point at infinity. Consider the gauge transform function defined

$$U(\mathbf{x}) = \exp\left(\frac{i\pi N \mathbf{x}\cdot\boldsymbol{\sigma}}{\sqrt{|\mathbf{x}|^2+R^2}}\right)$$

where $\mathbf{x}$ is a point in $\mathbb{R}^3$, and $\boldsymbol{\sigma}$ represents the Pauli matrices with R an arbitrary length scale. Show the winding number Eq. 5.15 gives the integer N. Why does N need to be an integer here?

Short Digression on Quantum Gravity[1]

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

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

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[2] experiment has ever probed. We must surely accept the 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.

6.0.2 Which Approach?

There are several approaches to quantum gravity. While I will not make any statement about which approaches are promising, and which approaches are crazy and overpublicized[3], I am comfortable stating that many of these investigations have led to incredibly 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.

6.1 Some General Principles?

We have to choose general principles that we 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

[1] This chapter aims to give context about why people first started studying topoogical theories. It can be skipped on a first reading (but do come back later to enjoy it!).

[2] 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.

[3] For some basic 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 (with appropriate warning that it a bit of a rant) the online response by Lubos Motl. Also enlightening are the online letters between Smolin and Lenny Susskind.

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 in all possible histories of the universe — it is the ultimate sum over 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[4] for gravity takes the form

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

where the integration is over the entire space-time manifold $\mathcal{M}$, where here g is the space-time metric tensor and R is the Ricci scalar[5]. 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

$$Z = \int \mathcal{D}g(x)\ \ e^{iS_{Einstein}[g(x)]/\hbar} \quad , \tag{6.2}$$

thus summing (or integrating) over all possible space-time metrics. Even without matter in the universe, the model is very nontrivial because the space-time metric can fluctuate — these fluctuations are just gravity waves[6]. Even in this limit no one has fully made sense of this type of path integral without many additional assumptions.

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

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 about this fact see the appendix to this 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 Λ in which case $R = 2\Lambda g$ is the solution).

One might think that this means that gravity in 2+1D is completely trivial. However, it is not. The space-time manifold, although every-

[4] Written down first by Hilbert in 1915.

[5] If you are rusty with your general relativity, recall that the metric tensor g defines the relativistically invariant line element via $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$, and the Ricci scalar R, which is a complicated function of g, is a measure of the curvature of a manifold which compares the volume a small ball to the volume it would have in flat Euclidean space. In particular for a D dimensional manifold $\mathcal{M}$ we would consider a D-dimensional ball B^D of radius ϵ and we have

$$\frac{V(B^D) \subset \mathcal{M}}{V(B^D) \subset \mathbb{R}^D} = 1 - \frac{\epsilon^2 R}{6(D+2)} + \dots .$$

[6] Observation of gravity waves by the LIGO experiment won the 2017 Nobel Prize. Long before this we had very strong indirect observation of gravity waves from observation of the Hulse-Taylor binary pulsar which earned a Nobel Prize in 1993.

where 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

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

where $S[g(x)]$ 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 analogous to that shown in Fig. 6.1.

Here is the surprise: the function $Z(\mathcal{M})$ is precisely the Chern-Simons partition function discussed above in section 5.2 for an appropriately chosen gauge group![7] This connection is very roughly sketched in the chapter appendix section 6.3.

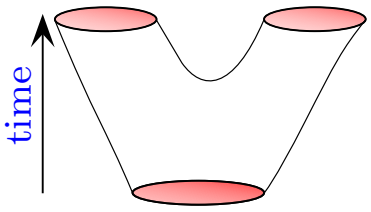


Fig. 6.1 A manifold where the topology of a space-like slice (slice at fixed time) changes as time progresses.

[7] This was first noted by Achúcarro and Townsend [1986] and then was developed further by Witten [1988] and many others.

6.1.1 Further Comments on Connections to Quantum Gravity

In the "this is not string-theory" school of thought for quantum gravity, evaluation of Eq. 6.2 is the main goal. Crucially one needs some variables to describe the metric of the universe. Several different approaches to this seem to converge on some similar structures. One interesting approach, known as loop quantum gravity, uses Wilson loop operators as the elementary variables of the theory (once one has reformulated gravity to look like a gauge theory). Another approach discretizes space-time and sums over the different possible discretizations[8]. With certain assumptions these approaches appear to be very closely related! In section 21.3 we will return to the issue of discretizing space-time and how this can result in topological gravity.

[8] Indeed at length scales as small as the Planck length $l_{Planck} = \sqrt{\hbar G/c^3} = \hbar c/E_{\text{Planck}} \approx 1.6 \times 10^{-35}$m, there is no reason to believe space-time resembles our macroscopic idea of a smooth manifold. The ratio of the radius of the sun to the radius of an atom is roughly the same as the ratio of the radius of an atom to the Planck length!

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 Carlip [2005])

> 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 $g = \eta + h$, 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

$$\epsilon_{\mu\nu}k^\nu = 0. \tag{6.3}$$

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. 6.3 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+1 D, but Eq. 6.3 is 3 constraints, and A has three parameters, so we should always be able to solve the equation for A given ϵ.

6.3 Appendix: Relation of 2+1D GR to Chern-Simons Theory (In Brief)

Let us start with a Chern-Simons Lagrangian for $SU(2)_k \otimes SU(2)_{-k}$. Here we will use a very shorthand notation

$$\mathcal{L} = \frac{k}{4\pi}\int_M (A_+ dA_+ + \frac{2}{3}A_+^3) + \frac{-k}{4\pi}\int_M (A_- dA_- + \frac{2}{3}A_-^3)$$

Making the transformation

$$\omega_= \frac{1}{2}(A_+ + A_-) \qquad e = \frac{k}{8\pi}(A_+ - A_-)$$

one obtains the Lagrangian (using differential form notation)

$$\mathcal{L} = \int (e \wedge R + \frac{\lambda}{3} e \wedge e \wedge e) \tag{6.4}$$

Here e is interpreted as the dreibein of general relativity which is related to the metric by (returning appropriate indices to vectors)

$$g_{\mu\nu} = e^a_\mu e^a_\nu \eta_{ab}$$

with η_{ab} the flat metric in 2+1 D, and ω is a spin connection which has an equation of motion that dictates it is torsion free, and the remaining Lagrangian Eq. 6.4 is precisely the 2+1D Einstein-Hilbert Lagrangian in the so-called Palitini form. In that equation

$$\lambda = (4\pi/k)^2$$

is the cosmological constant. The calculation here has been given for a Euclidean form of gravity. For Lorenzian gravity one needs to work with $SO(2,1)$ Chern-Simons theory which is a bit more complicated.

More details of the relationship between 2+1D general relativity and Chern-Simons theory are provided in the further reading, listed below.

Further Reading

- For a huge amount of information on 2+1 dimensional quantum gravity, see Carlip [2005].
- The relationship of 2+1 D gravity to Chern-Simons theory was first developed by Ana Achúcarro and Paul Townsend ([Achúcarro and Townsend, 1986])
- The relationship was further developed by Edward Witten (Witten [1988])
- Years later, the question was revisited by Witten [2007], where doubt is raised as to whether Chern-Simons theory is sufficient to fully describe gravity in 2+1 dimensions.
- A (potentially biased) history of various approaches to quantum gravity is given by Rovelli [2000].
- Reviews of loop quantum gravity are given by Rovelli [2008] and Nicolai et al. [2005].
- Discussions of discretization approaches to quantum gravity are given by Regge and Williams [2000] and Lorente [2006].
- The article by Nicolai and Peeters [2007] covers the connections between the loop and discretization approach fairly clearly.

Note that none of these references are particularly easy to digest!

Defining Topological Quantum Field Theory[1]

7

We already have a rough picture of a Topological Quantum Field Theory (TQFT) as a quantum theory that depends on topological properties as opposed to depending 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[2] 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[3]. 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 $Z(\mathcal{M}, \text{links})$ as we discussed in section 5.2. The advantage of thinking in this language is that strictly speaking, the functional integrals of Chern-Simons theory

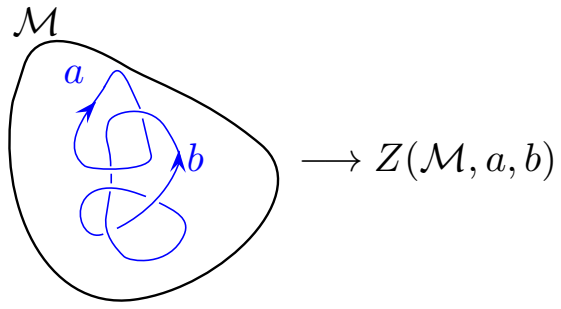


Fig. 7.1 A (2+1) dimensional TQFT takes an input of a labeled link in a manifold and produces an output of a complex number in a manner which is topologically invariant.

[1] Many students find this chapter frighteningly abstract. While this chapter sets the stage for a number of ideas that come later, it can also be skipped to a large extent if it seems too difficult. While it may seem a bit cruel to include such a chapter early in the book, I've included it here because it gives the best definition of what a TQFT actually is — which, in one form or another, is what we are studying for the remainder of the book.

[2] Particularly condensed matter physicists might start to wonder why we need to start talking about arbitrary, and potentially bizarre sounding, three dimensional manifolds — what could they possibly have to do with real physical systems? However (besides just being a beautiful digression) pursuing this direction allows us to understand some of the strong constraints on topological models and their mathematical structure, and this turns out to be important for the analysis of even fairly simple physical systems.

[3] We may also allow world lines of anyons to fuse into other species as discussed in section 4.2.

are often not well defined mathematically. Instead, here we bypass the Chern-Simons field theory altogether and define a TQFT simply as a mapping from a manifold with a link to an output.

A closely related but more formal definition of TQFTs is given by a set of Axioms by Atiyah [1988][4] which are in some sense much more informative.

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 in section 7.2 we will discuss adding particles and moving them around in space-time too.

We will consider a $D+1$ dimensional space-time manifold[5] which we call $\mathcal{M}$, and D dimensional oriented slice Σ — we can often think of this 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[6] 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 depend on the particular anyon theory we are considering. For example in the case of abelian anyons in section 4.3 we found a degeneracy of m for a system on a torus with statistical angle $\theta = \pi p/m$.

Note that when we add particles to the system (we will do this in section 7.2), if the particles are nonabelian, then there will also be a Hilbert space associated with the additional degeneracy that comes with such nonabelian 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[7]. 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] Sir Michael Atiyah, a Fields medalist, who went to primary school in Sudan, was one of the foremost mathematicians of the 20th century. He specialized in geometry and topology — particularly at the interface between mathematics and physics. You can find videos of him talking about life, physics, and mathematics at webofstories.com.

[5] 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. See section 33.1.

[6] The phrases "depends only on the topology..." is something that physicists would say, but mathematicians would not. To a mathematician, topology describes things like whether sets contain their limit points, whether points are infinitely dense and so forth. Perhaps it would be better to just say that $V(\Sigma)$ does not change under continuous deformation of Σ. This is something mathematicians and physicists would both agree on, and this is what we actually mean here!

[7] This may sound a bit abstract, but it is exactly how the Hilbert spaces of any two systems must combine together. For example, in the case of two spins, the Hilbert space of the union of the two spins is the tensor product of the two Hilbert spaces.

The reason this must be true is because $\emptyset \cup \Sigma = \Sigma$ and $\mathbb{C} \otimes V(\Sigma) = V(\Sigma)$ so the result follows[8].

AXIOM 3: If $\mathcal{M}$ is a $(d+1)$-dimensional manifold with D-dimensional boundary[9] $\Sigma = \partial M$, 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 might 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).

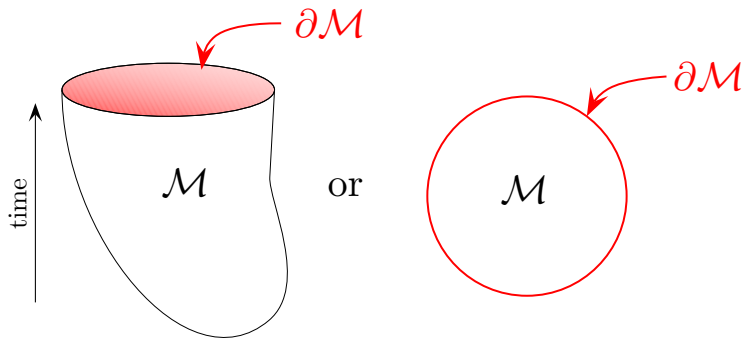


Fig. 7.2 Two depictions of a space-time manifold $\mathcal{M}$ with boundary $\partial \mathcal{M}$. The left depiction is problematic because the only boundary of the manifold is supposed to be the red top surface ∂M (the black outline of $\mathcal{M}$ really should not be there, but we can't draw a closed three manifold!). 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. This is exactly what we found from Chern-Simons theory.

[8] 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.

[9] We use the ∂ to denote boundary. See section 33.1.4.

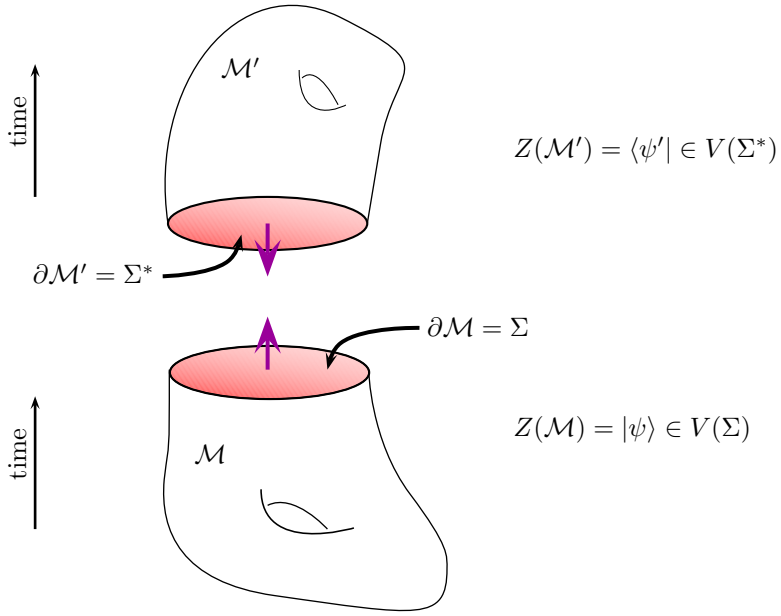


Fig. 7.3 In this picture $\mathcal{M}$ and $\mathcal{M}'$ are meant to fit together since they have a common boundary but with opposite orientation $\Sigma = \partial\mathcal{M} = \partial\mathcal{M}'^*$. Here $\langle\psi'| = Z(\mathcal{M}') \in V(\Sigma^*)$ lives in the dual space of $|\psi\rangle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

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 kets into bras. 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[10]

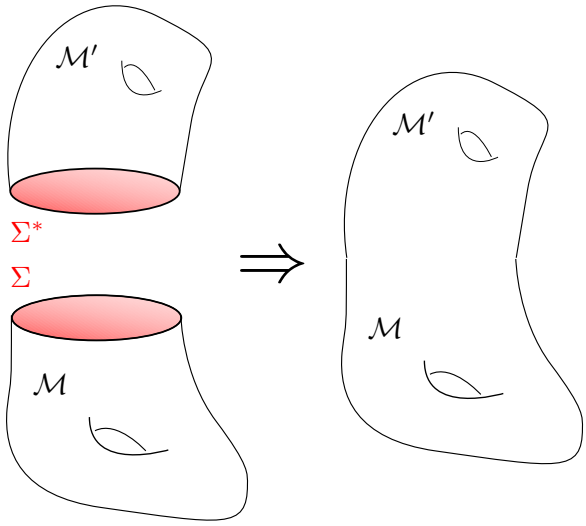
$$Z(\mathcal{M} \cup_\Sigma \mathcal{M}') = \langle Z(\mathcal{M}')|Z(\mathcal{M})\rangle \tag{7.1}$$

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

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



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

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

example.

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 an evolution[11] similar to that shown in Fig. 7.5.

[11] This evolution may or may not be unitary — indeed, the dimensions of $V(\Sigma_1)$ and $V(\Sigma_2)$ may not even match if $\Sigma_1 \neq \Sigma_2$.

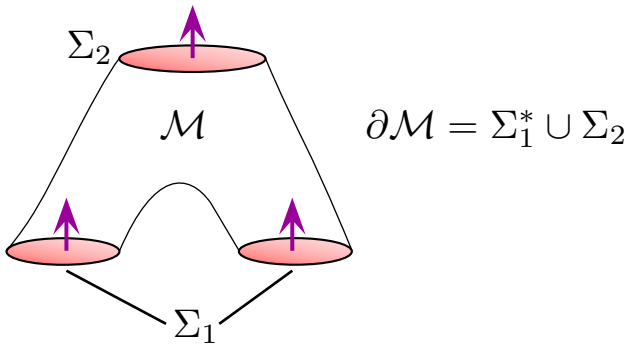
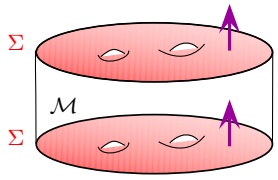


Fig. 7.5 $\mathcal{M}$ is the cobordism between Σ_1^* and Σ_2. I.e., $\partial\mathcal{M} = \Sigma_1^* \cup \Sigma_2$. Note that we have reversed orientation of Σ_1 here.

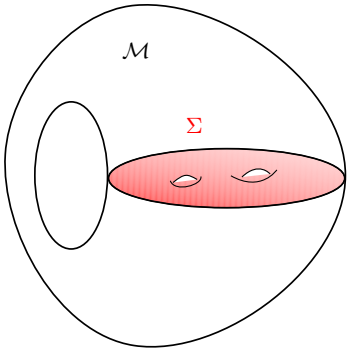
IDENTITY COBORDISM: If we have $\mathcal{M} = \Sigma \times I$ where I is the



$\mathcal{M} = \Sigma \times I$

$\partial\mathcal{M} = \Sigma \cup \Sigma^*$

Fig. 7.6 A cobordism that can be topologically contracted to nothing acts as the identity on the Hilbert space $V(\Sigma)$.



$\mathcal{M} = \Sigma \times S^1$

Fig. 7.7 Gluing the top of $\Sigma \times I$ to the bottom we obtain $\mathcal{M} = \Sigma \times S^1$. An important fact is that $Z(\Sigma \times S^1)$ is just the ground state degeneracy of the 2-manifold Σ

[12]For dimension $D > 2+1$ dimensional TQFTs we could have world-sheets of moving strings and other higher dimensional objects as well.

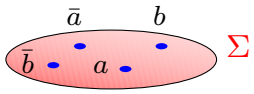


Fig. 7.8 A 2-manifold with particles in it, which are marked and labeled points. We now call the combination (the manifold and the marked points) Σ for brevity.

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 (or infinitesimal thickness"), we can take this map to be the identity:

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

where the sum is over the entire basis of states of $V(\Sigma)$.

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), as shown in Fig. 7.7. We then have

$$Z(\Sigma \times S^1) = \mathrm{Tr}\,[Z(\Sigma \times I)] = \mathrm{Dim}[V(\Sigma)]. \tag{7.2}$$

where Tr means trace. 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 4.3, for the torus T^2 we have

$$\mathrm{Dim}\, V(T^2) = \text{number of particle species} \tag{7.3}$$

which we argued (at least for modular abelian anyon models) based on non-commutativity of taking anyons around the handles of the torus, and we will justify for nonabelian anyons as well in section 7.2.1. Similarly, for a 2-sphere S^2, we have

$$\mathrm{Dim}\, V(S^2) = 1 \tag{7.4}$$

since there are no noncontractable loops, and this will also hold for both abelian and nonabelian theories. See section 4.3.1 for discussion of the ground state degeneracy of abelian theories on higher genus surfaces.

7.2 Adding Particles

We now consider extending the ideas of TQFT to space-time manifolds with particle world-lines in them.[12]

Let us imagine that there are different anyon types which we can label as a, b, c, and so forth. The corresponding antianyons are labeled with overbar $\bar{a}$, $\bar{b}$ and so forth as in section 4.2.2. We now imagine a 2-manifold with some marked and labeled points as shown in Fig. 7.8. We call the combination of the 2-manifold with the marked points Σ for brevity. As with the case without particles (AXIOM 1, in section 7.1), 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 nonabelian 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

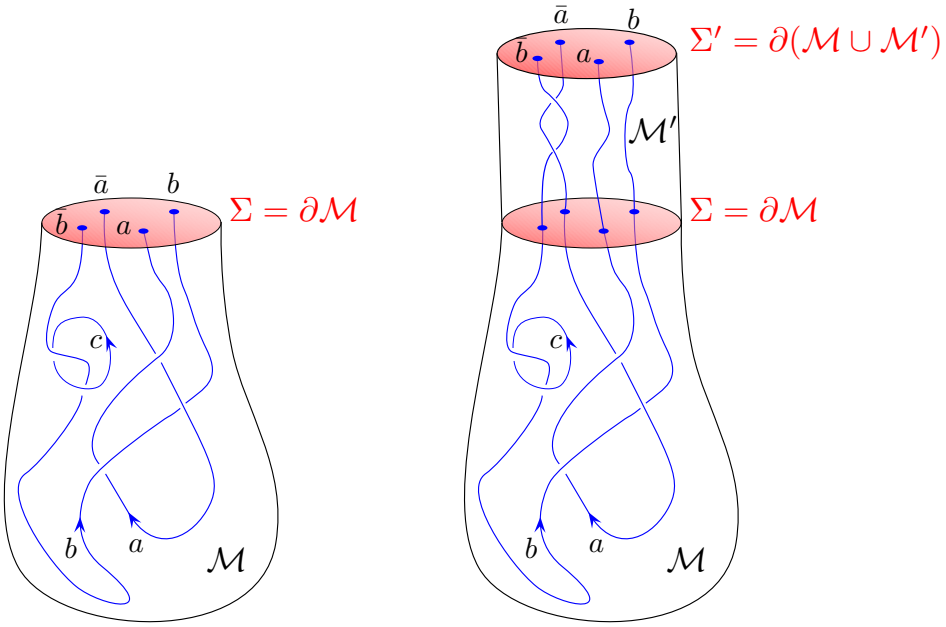


Fig. 7.9 Left: 3-manifold $\mathcal{M}$ with particles in it, which are marked and labeled lines (the lines should be directed unless the particle is its own antiparticle). These world lines may end on the boundary $\Sigma = \partial\mathcal{M}$. The wavefunction on the boundary $\partial\mathcal{M}$ is determined by the spacetime history given by $\mathcal{M}$. Right: $\mathcal{M}'$ 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. In this particular picture $\Sigma = \Sigma'$ being the same surface with the same types of particles at the same positions.

marked points around within the manifold (i.e., if we braid the particles around each other).

Similarly a 3-manifold $\mathcal{M}$ is now supplemented with labeled links indicating the world lines of the particles. The world-lines should be directed unless the particles are their own antiparticles. The world lines are allowed to end on the boundary of the manifold $\partial\mathcal{M}$. See left of Fig. 7.9. 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 discussion of axiom 3, the spacetime history specifies exactly which wavefunction

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

is realized on the boundary $\Sigma = \partial\mathcal{M}$. If a basis of $V(\partial M)$ is given by wavefunctions $|\psi_a\rangle$ then we can generally write the particular wavefunction $|\psi\rangle$ in this basis

$$|\psi\rangle = \sum_\alpha c_\alpha |\psi_\alpha\rangle.$$

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 the right of 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 write $|\psi'\rangle$ in the same basis as $|\psi\rangle$

$$|\psi'\rangle = \sum_\alpha c'_\alpha |\psi_\alpha\rangle.$$

We can then think of $Z(\mathcal{M}')$ as giving us a unitary transformation on this Hilbert space — which is exactly what we think of as nonabelian statistics. We can write explicitly the unitary transformation

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

or equivalently

$$c'_\alpha = \sum_\beta U^{\alpha\beta} c_\beta.$$

Note that if the particles stay fixed in their positions (or move in topologically trivial ways) then $\mathcal{M}'$ can be contracted to infinitesimal thickness and we can think of the unitary transformation as being the identity. As with the identity cobordism discussed in section 7.1, we can take such an identity transformation, glue the top to the bottom and obtain

$$Z(\Sigma \times S^1) = \mathrm{Dim}[V(\Sigma)] \tag{7.5}$$

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[13]. 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 tubular neighborhood 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 worm having eaten a path out of the manifold.) In this configuration, the boundary $\partial\mathcal{M}$ of the manifold $\mathcal{M}$ now contains the surface of these empty tubes — i.e, the surface of a torus T^2 for each world-line loop. Note that the empty tube is topologically a solid torus $D^2 \times S^1$ even if the world-line forms some knot[14]. The statement that it forms a nontrivial knot is a statement

[13] Up to here our discussion has been applicable to TQFTs in any dimension. From here on we specialize to the most interesting case of $D = 2$, that is 2+1 dimensions.

[14] D^2 is the usual notation for a two dimensional disk and S^1 again is the circle.

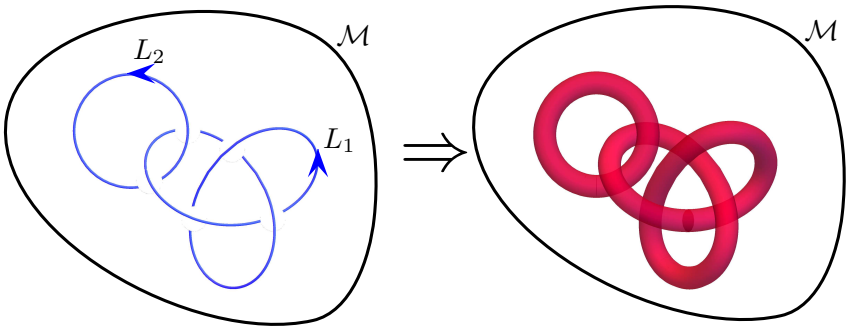


Fig. 7.10 Removing the world-lines on the left along with a thickened tube. Imagine a worm burrowing along the path of the world lines and leaving a hollow hole (colored red).

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. We can think of this solid torus 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. Obviously, gluing such a solid torus containing a particle world line (Fig. 7.11) back into the empty solid-torus-shaped tube (right of Fig. 7.10) recovers the original picture of labeled world lines following these paths (left of Fig. 7.10).

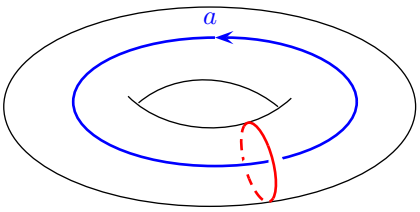


Fig. 7.11 The possible wavefunctions $|\psi_a\rangle$ that we can have on the surface of the torus can be realized by having a world-line of a particle of type a going around the handle of the torus. We can call these Z(solid torus with a running around handle) $= |\psi_a\rangle$

The partition function of the manifold with the tori excised from it (the right of Fig. 7.10) contains all of the information necessary to determine the 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 there are two surfaces (the two surfaces on the inside of the holes left where we excised the two tori), so we have

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

where $Z(\mathcal{M}; i, j)$ is the partition function for the torus with two particle types i, j following the two world line loops L_1 and L_2, and the two wavefunctions are the corresponding boundary condition. Thus, if we want to extract $Z(\mathcal{M}; a, b)$, where the particle lines are labeled with a, b we simply glue in the wavefunction $|\psi_{L1,a}\rangle \otimes |\psi_{L2,b}\rangle$ representing the boundary condition on the two surfaces.

7.3 Building Simple 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 using the gluing axiom from section 7.1. 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. See the discussion in section 33.1). 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[15]

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

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 B^4. Topologically we have[16]

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

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

$$\begin{aligned} S^3 = \partial B^4 &= \partial(D^2 \times D^2) = (\partial D^2 \times D^2) \cup (D^2 \times \partial D^2) \\ &= (S^1 \times D^2) \cup_{T^2} (D^2 \times S^1) \end{aligned}$$

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.

Let us think about the partition function of these two solid tori which

[15] If you are rusty on these elementary topology manipulations, see the review in section 33.1

[16] Topologically it is easiest to think about the n-dimensional ball, B^n, as being the interval $I = B^1$ raised to the n^{th} power. 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 $B^4 = I \times I \times I \times I = D^2 \times D^2$.

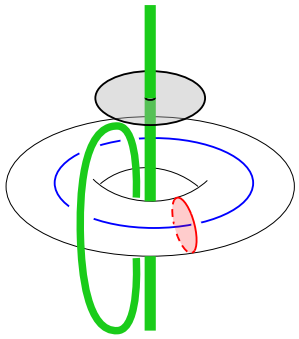


Fig. 7.12 Assembling two solid tori to make S^3. The obviously drawn torus $D^2 \times S^1$ can be thought of as the red disk D^2 crossed with the blue circle S^1. The remainder of space outside of this torus, including the point at infinity is the other solid torus $S^1 \times D^2$. For this "outside" solid torus, the S^1 can be thought of as the vertical green line. This line becomes S^1 by connecting up with itself at the point at infinity. The upper shaded disk is an example of a contractable D^2 which is contained entirely within the outside solid torus. Note that the entire outside solid torus is $S^1 \times D^2$, the vertical green line crossed with disks topologically equivalent to this one. The green loop off to the side (also contained within the outside torus), like the vertical green S^1 loop is not contractable within the outside solid torus, but can be deformed continuously to the vertical green loop.

are glued together on their boundaries to make up S^3. We write the partition function as the overlap between wavefunctions on the outside and inside tori:

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

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

$$Z(S^3; a \text{ loop linking } b \text{ loop}) = \langle Z(S^1 \times D^2; b) | Z(D^2 \times S^1; a) \rangle \equiv S_{ab} \tag{7.6}$$

This quantity S_{ab} is known as the **modular S-matrix**, and it is a very important quantity in topological theories as we shall see in chapter 17 below.[17]

[17] 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 $\bar{a}$ is the antiparticle of a.

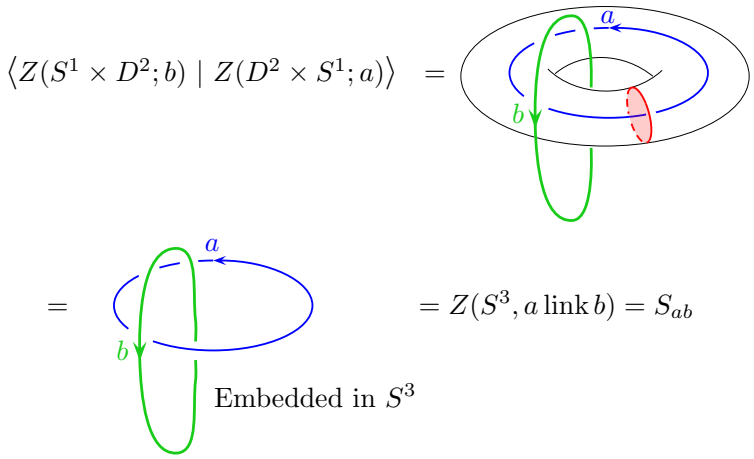


Fig. 7.13 Here we assemble a partition function for S^3 with world lines of a linking b embedded in the S^3. To do this we glue together two solid tori each with a world line running around the handle. The green line marked b runs around the handle of the "outside" torus. The end result is known as the 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.

Note that the S-matrix is unitary[18], since it is simply a basis transformation between the two sets of wavefunction which both span the vector space $V(T^2)$ of the torus surface T^2 where the two solid tori are glued together. Note also that the element S_{00}, corresponding to 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

[18] Here we are assuming the theory is *modular* meaning there are no transparent particles. This assumption will be discussed in more depth in section 17.3.

tells us that

$$Z(S^3) = S_{00} \leq 1 \tag{7.7}$$

and in fact, should be strictly less than one unless there are no nontrivial particle types and S is a one-by-one dimensional matrix.

Another way of viewing the S matrix is as a simple link between two strands, as shown in Fig. 7.13. As with the Kauffman bracket invariant, we can construct a set of diagrammatic rules to give a value to knots. Soon, in chapters 8-16 we will construct diagramatic rules to help us "evaluate" knots like this. These rules will be somewhat similar to the rules for the Kauffman bracket invariant, only now we need to keep track of labels on world lines as well.

7.3.2 $S^2 \times S^1$

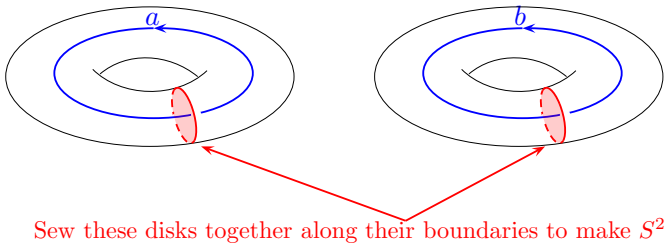


Fig. 7.14 Assembling two solid tori to make $S^2 \times S^1$. Here the two contractable disks D^2 are sewed together along their boundaries to make S^2.

There is another way we can put two solid tori together to make a closed manifold[19]. Instead of attaching longitude-to-meridian and meridian-to-longitude, we instead attach meridian-to-meridian and longitude-to-longitude. (This is perhaps a simpler way to put together two solid tori!) See Figure 7.14. Here we claim that[20]

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

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 along its equator and getting two disks which are the north and south hemispheres).

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[21]

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

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 tori.

[19] In fact there are an infinite number of ways two tori can be sewed together to form a closed manifold. These are discussed in detail in the appendix to this chapter, section 7.4.

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

[21] It is worth considering how the world lines, in the case where $a = b$, are positioned in the $S^2 \times S^1$. The world line around the handle of one torus enters each S^2 sphere through one hemisphere and the world line around the handle of the other torus exits each S^2 sphere through the other hemisphere. This fits with the principle that a nonzero amplitude of two particles on the surface of a sphere can only occur if the two particles are a particle-antiparticle pair. This is discussed in section 8.4.

In particular, we have the case where we put no particle (the vacuum) around both handles, we have (i.e., $a = b = I = 0$)

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

So we have the result

$$Z(S^2 \times S^1) = 1 \tag{7.8}$$

Note that this agrees with two of our prior statements. On the one hand Eq. 7.5 says that Z for any two dimensional manifold crossed with S^1 should be the dimension of the Hilbert space for that manifold; and on the other hand Eq. 7.4 states that the dimension of the Hilbert space on a sphere is 1.

7.4 Appendix: Sewing Two Solid Tori Together

While this discussion is a bit outside the main train of thought (being the development of TQFTs) it is interesting to think about the different ways two solid tori may be sewed together to obtain a closed manifold.

A solid torus is written as $D^2 \times S^1$. We define the meridian m to be the S^1 boundary of any D^2. I.e., the meridian is a loop on the surface around the contractable direction of the solid torus. We define the longitude l as being any loop around the surface of the solid torus which intersects a meridian at one point. This definition unfortunately has some (necessary) ambiguity. A line that loops around the meridian n times as it goes around the noncontractable direction of the torus, is just as good a definition of a longitude (an example of this is Fig. 7.15 which is $n = 1$). We call this line $l + nm$ where n is the number of times it goes around the meridian and l was the original definition of the longitude that did not loop around the meridian. Redefining the longitude this way is known as a "Dehn Twist".

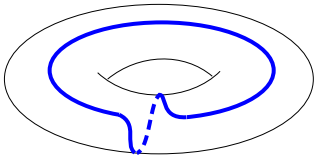


Fig. 7.15 A line that wraps both the longitude and meridian of the torus.

Let us choose a meridian m_1 on the surface of one solid torus and choose to sew it to the line $-qm_2 + pl_2$ of the second solid torus (that is, the line that goes p times around the longitude and $-q$ times around meridian, we make $-q$ negative so that the two tori surfaces are opposite oriented for attaching them together. Once the two lines are glued together this uniquely defines how the rest of the two torus surfaces are glued together. The resulting object is known as the "Lens space" $L(q, p)$. In section 7.3.1 we showed that $L(0, 1) = S^3$ and in section 7.3.2 we showed that $L(1, 0) = S^2 \times S^1$. Note that due to the ambiguity of definition of the longitude of the torus $-qm_2 + pl_2$, under redefinition of the longitude goes to $(-q - np)m_2 + pl_2$. Thus $L(q + np, p) = L(q, p)$, and in particular, $L(1, 1) = S^3$ also.

Chapter Summary

- The Atiyah Axioms formalize the idea of a topological quantum field theory.

Further Reading

For discussion on the Atiyah Axioms see Atiyah [1988, 1997]

A discussion of gluing together manifolds (as in sections 7.3-7.4) is given by Rolfson [1976]. The book Farb and Margalit [2012] may also be useful. We will discuss this type of sewing further in chapter 22 below.

Part II

Anyon Basics

Fusion and Structure of Hilbert Space

8

As discussed in section 7.1, 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. At the same time we will building up a diagrammatic algebra with the goal of constructing a mapping from world-lines of particles to complex numbers (a definitions of a TQFT as given in Fig. 7.1). We briefly introduced graphical notation in section 4.2.1 and we will continue that development here. For those who prefer more mathematical detail, in section 8.6 (as well as in chapter 12) we will introduce tensor description of diagrams and the associated Hilbert spaces.

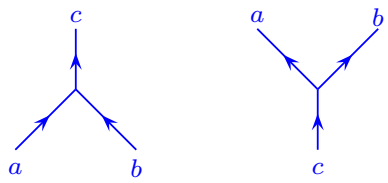


Fig. 8.1 Fusion (left) and splitting (right) diagrams can be thought of as part of a space-time history of the particles. If we are describing two separated particles a and b whose overall quantum number is c (sometimes we say "overall fusion channel is c), we would describe the ket for this state using the right hand picture — which we can think of as a space-time description of how the current situation (a on the left b on the right) came about (with time going up). Details of the formal meaning of these diagrams in terms of as bras and kets is given in sections 8.6 and chapter 12.

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 $a \times b = b \times a = c$. Fusion and splitting are shown diagrammatically in Fig. 8.1. Sometimes colloquially we call both diagrams "fusion".

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.

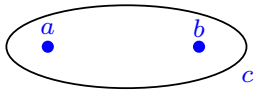


Fig. 8.2 Another notation to describe the fusion of two particle types to make a third $a \times b = c$. The two particles need not be close to each other. This figure is equivalent to the right of Fig. 8.1.

In our abelian anyon model of charges and fluxes (see section 4.2), 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

$$a \times I = a$$

for any particle type a. In the charge-flux model (section 4.2) we should think of the identity as being no charge and no flux. Fusion with the identity is depicted schematically in Fig. 8.3. Often we do not draw the identity particle at all, being that it is equivalent to the absence of any (nontrivial) particle.

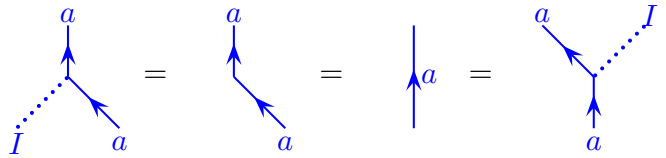


Fig. 8.3 Fusion of a particle with the identity $a \times I = a$. The dotted line indicates the identity. In some of these pictures the a particle appears to move slighly to the left. However, this is not important for topological properties since the path can be deformed continuously to a particle that does not move.

[1] It is annoying that we have so many different ways to express the identity, but in different contexts different notations seem natural. For example, if our set 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. See note 5 in chapter 33.

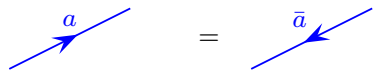


Fig. 8.4 A particle going forward should be equivalent to an antiparticle going backwards.

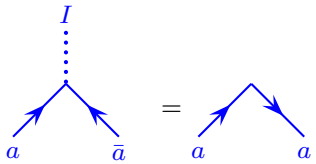


Fig. 8.5 Fusion of an anyon with its anti-anyon to form the identity can be thought of as a particle turning around in space-time. On the right, we have reversed one arrow and changed $\bar{a}$ to a, and we have not drawn the identity line.

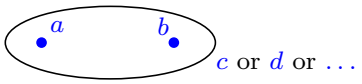


Fig. 8.6 Multiple possible fusion channels. Here we show that a and b can fuse together to give either c or d or other possible results.

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 in time should be equivalent to an antiparticle going backwards in time as shown in Fig. 8.4. Fusion to the identity can be thought of as a 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 from section 4.2 would be the "2" particle (fusion of 2 elementary anyons, see section 4.3) in the case of $\theta = \pi/4$. Also, the identity particle I is always its own antiparticle.

8.2 Multiple Fusion Channels - the Nonabelian Case

For the nonabelian theories as we have discussed above (for example in Section 3.5), 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 orthogonal fusion channels

$$a \times b = c + d + \ldots \tag{8.1}$$

meaning that a and b can come together to form either c or d or ..., as shown in 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 of Eq. 8.1).

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.

If two particles (maybe far away from each other) fuse together to some overall 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 d. 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

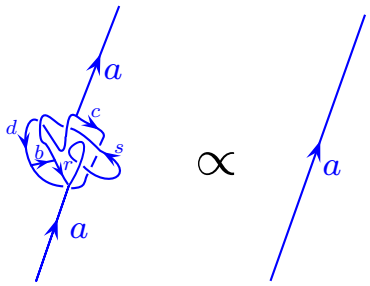


Fig. 8.7 If a particle a goes into a spacetime region, then a net particle charge a must come out. This is also sometimes called the "no-transmutation" principle. From far away, one can ignore any local processes (up to an overall constant).

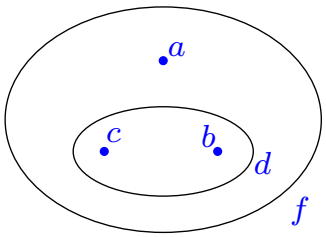


Fig. 8.8 In this picture b and c fuse to d. Then this d fuses with a to give an overall fusion channel of f. No amount of braiding b around c will change the fact that the two of them fuse to d. However, if we braid a with b and c, this can change the fusion of b with c subject to the constraint that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

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.

Locality gives another important way in which of anyons differs from the fusion of spins. With spins, if you can measure two spins individually you can (at least sometimes) determine the fusion channel of the spins. For anyons you must be able to measure a loop that *surrounds* both anyons in order to determine their collective fu sion channel — measuring each anyon individually does not tell you the fusion of the two!

Antiparticles in the Case of Multiple Fusion Channels

When we have multiple fusion channels (i.e., for nonabelian theories) we define antiparticles via the principle that a particle *can* fuse with its antiparticle to give the identity, although other fusion channels may be possible.

$$a \times \bar{a} = I + \text{ other fusion channels}$$

As in the abelian case we use the overbar notation to indicate an antiparticle. It should be the case that for each particle a there is a unique particle that can fuse with it to give the identity, and we call this particle $\bar{a}$. As in the abelian case, a particle may be its own antiparticle if $a \times a = I +$ other fusion channels, in which case we do not put an arrow on the line corresponding to the particle.

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 which we will study in more depth in section 31. Fibonacci anyons are closely related to the $SU(2)_3$ Chern-Simons theory[3].

In this example 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

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

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 nontrivial. This final rule also 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, as shown in Fig. 8.9.

With three Fibonacci anyons the Hilbert space is 3 dimensional, as

[2] Fibonacci, also known as Leonardo of Pisa, was born around 1175 AD. Perhaps his most important contribution to mathematics is that he 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!

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

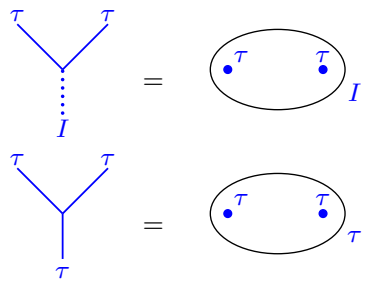


Fig. 8.9 Two different notations for the two different fusion channels of two Fibonacci anyons

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.

There is a single state in the Hilbert space of three anyons with overall fusion channel I. This state is labeled as[4] $|N\rangle$. As mentioned above by Fig. 8.7, due to locality, no amount of braiding amongst the three particles will change this overall fusion channel (although braiding may introduce an overall phase).

There are two states in the Hilbert space of three anyons with overall fusion channel τ. These are labeled $|1\rangle$ and $|0\rangle$ in Fig. 8.10. Again, as mentioned above by Fig. 8.7, 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 from $|0\rangle$ to $|1\rangle$ or from $|1\rangle$ to $|0\rangle$. However, as we will see below in section 10.1, 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 $|1\rangle$.

[4]Here $|N\rangle$ stands for "noncomputational", since it is not used in many quantum computing protocols that use Fibonacci anyons.

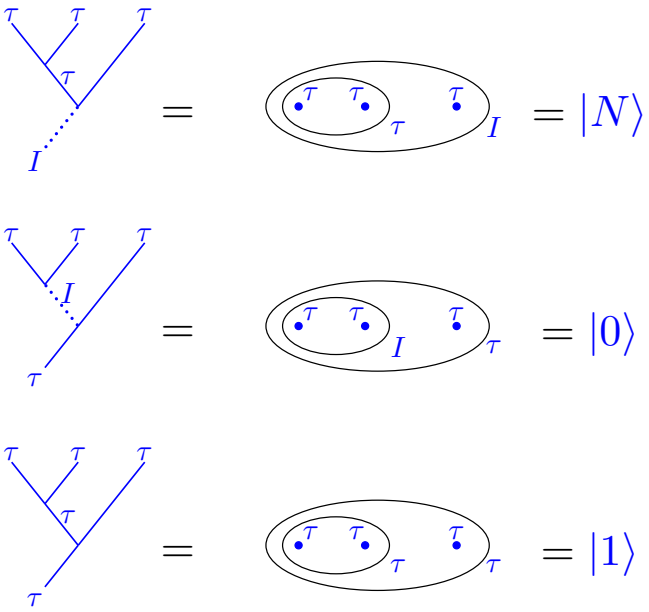


Fig. 8.10 Notations for the three different orthogonal fusion channels of three Fibonacci anyons. The notation $|N\rangle, |1\rangle$ and $|0\rangle$ are common notations for those interested in topological quantum computing with Fibonacci anyons!

For our Fibonacci system, with 2 particles the Hilbert space is 2 dimensional. With 3 particles the Hilbert space is 3 dimensional. It is easy to see that with 4 particles the Hilbert space is 5 dimensional (fusing a fourth anyon with $|0\rangle$ or $|1\rangle$ in figure 8.10 can give either I or τ, whereas fusing a fourth anyon with $|N\rangle$ can only give τ, thus giving a space of

dimension 2+2+1). With five particles the space is 8 dimensional and so forth. This pattern continues following the Fibonacci sequence (Try to show this!), hence the name.

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

$$\text{Dim of } N \text{ Anyons} = \text{Fib}_N \sim \left(\frac{1+\sqrt{5}}{2}\right)^N \tag{8.2}$$

We say that the *quantum dimension* of this particle is $\mathsf{d} = (1+\sqrt{5})/2$, the golden mean (See Eq. 3.8).

8.2.2 Example: Ising Anyons

The Ising[5] anyon system is extremely closely related to $SU(2)_2$ Chern-Simons theory[6], and this general class of anyon is believed to be realized in the $\nu = 5/2$ quantum Hall state (see section 31), topological superconductors, and other so-called Majorana systems (see section ***).

The Ising theory has three particle types[7]:

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

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 anything with the ψ particle always gives a unique result on the right hand side. We thus call ψ an abelian particle (despite the fact that the full theory is nonabelian), or we say that ψ is a *simple current* (see margin note 8 below). Fusion of many ψ particles is therefore fairly trivial, since each pair fuses to the identity in only one way.

Fusion of many σ particles, however, 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. Such a fusion tree is shown in Fig 8.11. By counting possible trees, we find that the total number of different fusion channels for N particles of type σ is $2^{N/2}$ (rounding down if $N/2$ is not an integer). 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 two state system, or 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 $\mathsf{d} = \sqrt{2}$ (See Eq. 3.8).

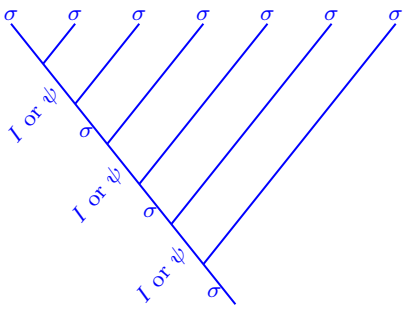


Fig. 8.11 The fusion tree for many σ particles in the Ising anyon theory.

[5] The name Ising is used here due to the relationship with the Ising conformal field theory which describes the Ising model in 2D at its critical point.

[6] The fusion rules of Ising and $SU(2)_2$ are the same, but there are some spin factors which differ, as well as a Frobenius-Schur indicator — see section ***.

[7] Another common notation is to use ϵ instead of ψ in the Ising theory. In $SU(2)_2$ the particles I, σ, ψ may be called $0, 1/2, 1$ or $0, 1, 2$.

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 antiparticle.

The general fusion rules can be written as

$$a \times b = \sum_c N_{ab}^c \; c \tag{8.3}$$

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[8] .

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[9] in which case a diagram with a vertex showing a and b fusing to c should also contain an index ($\mu \in 1 \ldots N_{ab}^c$) at the vertex indicating which of the possible c fusion channels occurs, as shown in Fig. 8.12. For most simple anyon theories N_{ab}^c is either 0 or 1, and we will not usually consider the more complicated case in examples for simplicity, but they are discussed in the chapter appendices for completeness (See section 9.5.3. See also section ***). It is good to keep in mind that such more complicated cases exist.

[8] A particle a is a *simple current* if $\sum_c N_{ab}^c = 1$ for each particle b.

[9] 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 fusion of representations of $SU(3)$. Recall that

$$8 \otimes 8 = 1 \oplus 8 \oplus 8 \oplus 10 \oplus \bar{1}0 \oplus 27$$

and the 8 occurs twice on the right.

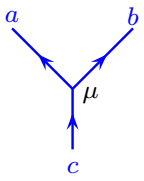


Fig. 8.12 Multiple fusion channels. In nonabelian theory fusion of a and b to c can occur in multiple orthogonal ways when $N_{ab}^c > 1$. To specify which way they fuse, we add an additional index $\mu \in 1 \ldots N_{ab}^c$ at the vertex as shown.

Elementary properties of the fusion multiplicity matrices

- Commutativity of fusion $a \times b = b \times a$.

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

- Time reversal

$$N_{ab}^c = N_{\bar{a}\bar{b}}^{\bar{c}} \tag{8.4}$$

- Trivial fusion with the identity

$$N_{aI}^b = \delta_{ab} \tag{8.5}$$

- Uniqueness of inverse

$$N_{ab}^I = \delta_{b\bar{a}} \tag{8.6}$$

It is sometimes convenient to define

$$N_{ab\bar{c}} = N_{ab}^c \tag{8.7}$$

which is the number of different ways that a, b, and $\bar{c}$ can fuse to the identity. An example of this equivalence is shown graphically in Fig. 8.13. The advantage of this representation is that N_{abc} is fully symmetric in all of its indices. For example, using this notation Eq. 8.5 and Eq. 8.6 are actually the same. Further, using Eq. 8.7 along with the symmetry

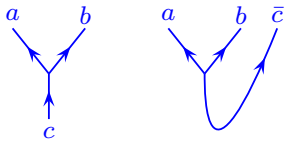


Fig. 8.13 An equivalence of N_{ab}^c with $N_{ab\bar{c}}$. Both types of vertices have the equivalent fusion multiplicity. Note that the left half of the right picture is exactly equivalent to the right — c is entering the vertex from below (then this c turns over to become a $\bar{c}$ going up on the far right).

of N_{abc} we can derive identities such as

$$N_{ab}^{c} = N_{a\bar{c}}^{\bar{b}} = N_{\bar{a}b}^{c}. \tag{8.8}$$

where in the last step we used Eq. 8.4.

Fusing Multiple Anyons

If we are to fuse, say, five anyons of type a together into a final result of e, we can do so via a tree as shown in Fig. 8.14.

To find the dimension of the Hilbert space, we write

$$\begin{aligned} \text{Dim of fusing five } a \text{ anyons to final result } e &= \sum_{bcd} N_{aa}^{b} N_{ba}^{c} N_{ca}^{d} N_{da}^{e} \\ &= \sum_{bcd} N_{aa}^{b} N_{ab}^{c} N_{ac}^{d} N_{ad}^{e} \end{aligned}$$

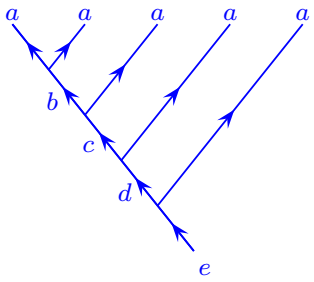


Fig. 8.14 Fusing five a type anyons together into a final result e.

and we identify each factor of N as being one of the vertices in the figure. It is convenient to think of the tensor N_{ab}^{c} as a matrix N_a with indices b and c, i.e, we write $[N_a]_b^c$, such that we have

$$\text{Dim of fusing five } a \text{ anyons to final result } e = [(N_a)^4]_a^e$$

Similarly were we to have a larger number p of anyons of type a we would need to calculate $[N_a]^{p-1}$. We recall (See Eq. 3.8) that the quantum dimension d_a of the anyon 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. We thus have that

$$\mathsf{d}_a = \text{largest eigenvalue of } [N_a] \tag{8.9}$$

Note that this implies $\mathsf{d}_a = \mathsf{d}_{\bar{a}}$ given the symmetries of N.

Example of Fibonacci Anyons

The fusion matrix for the τ particle in the Fibonacci theory is

$$N_\tau = \begin{array}{c} \begin{array}{cc} I & \tau \end{array} \\ \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{array}{c} I \\ \tau \end{array} \end{array}$$

where, as indicated 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 the 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 $\mathsf{d}_\tau = (1+\sqrt{5})/2$, in agreement with Eq. 8.2.

Example of Ising Anyons

The fusion matrix for the σ particle in the Fibonacci theory is

$$N_\sigma = \begin{array}{c} \begin{array}{ccc} I & \sigma & \psi \end{array} \\ \left(\begin{array}{ccc} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{array} \right) \end{array} \begin{array}{c} I \\ \sigma \\ \psi \end{array}$$

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 $\mathsf{d}_\sigma = \sqrt{2}$ as described in section 8.2.2.

8.3.1 Associativity

It should be noted that the fusion multiplicity matrices N are very 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. 8.15 or we can try fusing b and c first as on the right. Whichever we choose, we are describing the same Hilbert space and we should find the same overall dimension either way. In other words, we should have the same total number of fusion channels. Thus, corresponding to these two possibilities we have the equality

$$\sum_d N_{ab}^d N_{cd}^e = \sum_f N_{cb}^f N_{af}^e \tag{8.10}$$

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

$$[N_a, N_c] = 0 \tag{8.11}$$

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 (Since $[N_a, N_{\bar{a}}] = 0$ and $N_a = N_{\bar{a}}^T$ by Eq. 8.8). A set of normal matrices that all commute can be simultaneously diagonalized, thus

$$[U^\dagger N_a U]_{xy} = \delta_{xy} \lambda_x^{(a)} \tag{8.12}$$

and all N_a's get diagonalized with the same unitary matrix U. Surprisingly (as we will see below in section 17.3.1) for well behaved (so-called "modular"[10] anyon theories) the matrix U is precisely the modular S-matrix we discussed above in Eq. 7.6 !

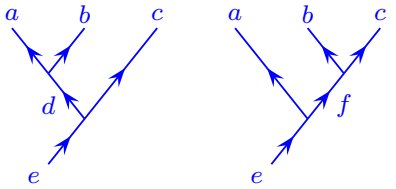


Fig. 8.15 Fusing $(a \times b) \times c$ should be equivalent to $a \times (b \times c)$. On the left a and b fuse to d first then this composite fuses with c to give e. On the right b and c fuse to f first, then this composite fuses with a to give e. Both diagrams represent the same physical Hilbert space. Fixing a, b, c, e the figure on the left spans the Hilbert space with different values of d whereas the figure on the right spans the same space with different values of f.

[10] For nonmodular theories, we can still diagonalize N in the form of Eq. 8.12, and the resulting unitary matrix U is sometimes known as the *mock* S-matrix.

8.4 Application of Fusion: Dimension of Hilbert Space on 2-Manifolds

The structure of fusion rules can be used to calculate the ground state degeneracy of wavefunctions on 2-dimensional manifolds[11]. Here we will again be examining the Hilbert space $V(\Sigma)$ where Σ is our 2-manifold which may or may not have particles in it.

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 in Eq. 7.5, there is a unique ground state in this situation because there are no non-contractable loops (See sections 7.1 and 4.3.1). 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 (nontrivial[12]) anyon on the sphere. In fact such a thing is not possible because you can only create particles in a way that conserves that overall quantum number. If we start with no particles on the sphere, 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

$$\text{Dim } V(S^2 \text{ with one (nontrivial) anyon}) = 0 \tag{8.13}$$

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 (non-vacuum) particle on the sphere.[13]

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[14]

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

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: the fusion of all the particles on the sphere must be the vacuum, since these particles must have (at some point in history) been pulled from the vacuum.

Now we could also imagine puncturing the sphere to make a hole where the particles were. In the spirit of what we did in section 7.2.1 we could re-fill the hole with any particle type[15] . 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

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

[11] We are again assuming manifolds are always orientable – so this excludes objects like the Klein bottle or the Möbius strip. Only a subset of TQFTs are well defined in the non-orientable case.

[12] By nontrivial we mean this particle is not the vacuum particle.

[13] For higher genus surfaces with non-abelian theories it is possible to have a single anyon alone on the surface. An example of this is when $a \times \bar{a} = I + c$. In this case a pair a and $\bar{a}$ may be created, one particle can move all the way around a handle to fuse with its partner, but it may leave behind a single anyon c since some quantum numbers can be changed by the action of moving the anyon around the handle. If we try this on the sphere (without the handle) we would always find that the pair re-annihilates to the vacuum. See further discussion near Eq. 8.14.

[14] It is implied that we are counting states here with the particles a and $\bar{a}$ at some given fixed position (all positions being topologically equivalent). If we were to count different positions as different states in the Hilbert space we would have to include this nontopological degeneracy in our counting as well.

[15] Since there is a time direction S^1_{time} as well, removing a disk with a particle in it from a spatial manifold Σ is precisely the same as removing a tubular neighborhood with a particle world line in it from the space-time manifold.

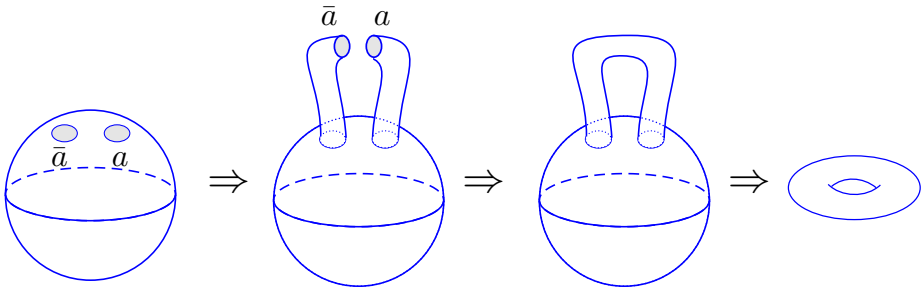


Fig. 8.16 Surgering the twice punctured sphere into a torus. This is the gluing axiom in action. Note that we are implicitly assuming the system is trivial in the "time" direction, which we assume to form a circle S^1_{time}.

Now consider the procedure shown in Fig. 8.16. 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 sometimes called *surgery* (We will discuss surgery in more detail in chapter 22). The result of this surgery is the torus surface T^2 and we conclude that

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

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 property of the TQFT. Although we are gluing together pieces along a 1-dimensional boundary (the edge of the punctures), we should realize that there is also a time direction, which we have implicitly assumed is compactified into S^1_{time}. Thus we are actually 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 wavefunctions on these two-surfaces ensures that the quantum number on these two punctures are conjugate to each other[16].

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. 8.17. We can then think of the sphere with three particles as being a sphere with three labeled punctures which is known as a "pair of pants", for reasons that are obvious in Figure. 8.18. 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 — this is known as a "pants decomposition". For example, in Fig. 8.19 we sew together two pair of pants to obtain a two handled torus.

[16] In Eq. 7.3 we had a torus surface which we crossed with an interval of time and we closed up the interval to form a circle, thus giving $\text{Tr}[Z(T^2 \times I_{time})] = Z(T^2 \times S^1_{time}) = \text{Dim}\, V(T^2)$. In contrast, in Fig. 8.16 we have a cylinder $S^1 \times I$ (topologically the same as a sphere with two holes) crossed with S^1_{time} and we close the cylinder to get $\text{Tr}[Z((S^1 \times I) \times S^1_{time})] = Z(T^2 \times S^1_{time})$.

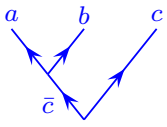


Fig. 8.17 Three particles that fuse to the identity. There are $N_{abc} = N^{\bar{c}}_{ab}$ different fusion channels.

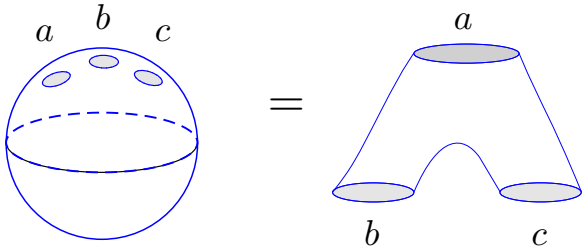


Fig. 8.18 A three-times punctured sphere is known as a "pair of pants".

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 assemble the manifold using two pair of pants as shown in Fig. 8.19 and then we simply need to figure out the number of possible fusion channels where we could satisfy $a \times b \times c \to I$ (for the bottom pair of pants) and $\bar{a} \times \bar{b} \times \bar{c} \to I$ (for the top pair of pants). This number of possible fusion channels is given in terms of the fusion multiplicities N_{abc} as shown in Fig. 8.17. Essentially we are just looking at the number of ways we can assign labels to the punctures when we glue the objects together. Thus we have

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

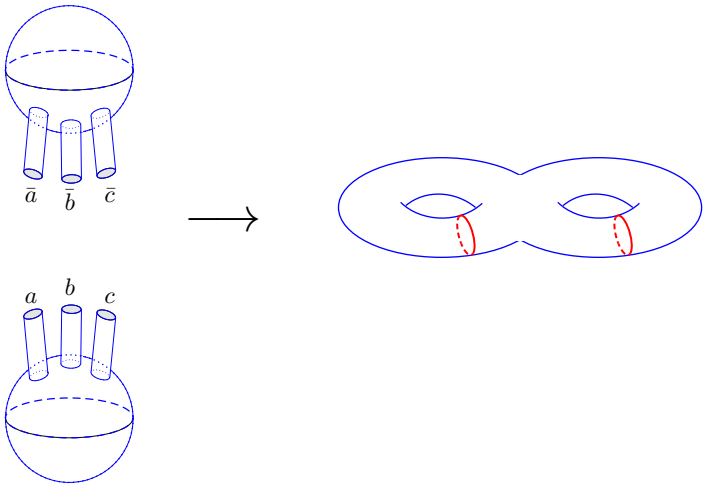


Fig. 8.19 Sewing together two pair of pants to form a two-handled torus.

Another interesting use of the pants diagram is to determine the degeneracy of a torus T^2 with a single anyon on it labelled a. Unlike the sphere, where one cannot have a single anyon on the surface (See Eq. 8.13) one can have a single anyon on a torus (See note 13 of this chapter). To see how this is possible, take a pants diagram with the

holes labelled $b, \bar{b}$, and a. Connect up the b to the $\bar{b}$ to give a torus with a single puncture remaining labeled a. Thus we conclude that

$$\text{Dim } V(T^2 \text{ with one } a) = \sum_b N_{b\bar{b}a} \equiv L_a \tag{8.14}$$

where we have defined this quantity to be called L_a.

One final example is to determine the ground state degeneracy of a three handled torus. There are many ways we might cut a three handled torus into pieces, but a convenient decomposition is the one shown in Fig. 8.20. Here there are three tori each with a puncture in it (marked as a red collar), and a single pants in the middle connecting the three. Each torus with a puncture has a Hilbert space dimension L_a where a is the quantum number assigned to the puncture. Thus the total dimension of the Hilbert space is conveniently written as

$$\text{Dim } V(\text{Three handled Torus}) = \sum_{abc} L_a L_b L_c N_{\bar{a}\bar{b}\bar{c}} \tag{8.15}$$

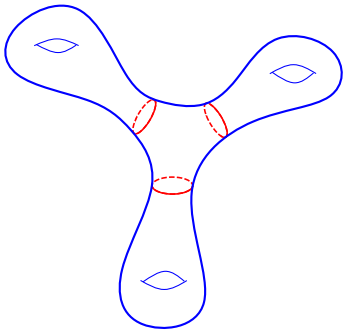


Fig. 8.20 Decomposing a three handled torus into three copies of a torus with puncture (the puncture is the red collar), and a single pants in the middle. I have resisted the urge to draw a three handled object as being covered with moss.

Example: Fibonacci Anyons

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

$$\begin{aligned} N_{III} &= 1 \\ N_{\tau\tau I} = N_{\tau I\tau} = N_{I\tau\tau} &= 1 \\ N_{\tau\tau\tau} &= 1 \end{aligned}$$

and all other $N_{abc} = 0$. Thus there are five possible labelings of the punctures in a pants diagram that allow overall fusion to the identity. If we match these together on both top and bottom of the diagram on the left of Fig. 8.19, we conclude that in the Fibonacci theory we have

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

Similarly, we can consider the degeneracy of states for a torus with a single τ particle on its surface

$$\text{Dim } V(T^2 \text{ with one } \tau \text{ particle on it}) = 1$$

coming from the allowed fusion $N_{\tau\tau\tau} = 1$. Thus we have $L_I = 2$ and $L_\tau = 1$. It is then easy to plug into Eq. 8.15 to obtain

$$\text{Dim } V(\text{Three handled torus}) = 15.$$

8.5 Product Theories

A very common construction is to consider the product of two anyon theories. Given two anyon theories (let us call them T and t) with

particle types

$$\begin{aligned} a, b, c, \ldots &\in t \\ A, B, C, \ldots &\in T \end{aligned}$$

we consider the product theory $T \times t$. The Hilbert space of the product theory is just the product of the Hilbert spaces of the constituent theories. So any arbitrary particle α in the product theory is composed of one particle from each of the consitutent theories

$$\alpha \in T \times t \qquad \Longrightarrow \qquad \alpha = (Y, x) \text{ with } Y \in T \text{ and } x \in t$$

For example, in the theory (Ising $\times$ Fibonacci), there are 6 particle types which we can label as

$$(I, I) \quad (I, \tau) \quad (\sigma, I) \quad (\sigma, \tau) \quad (\psi, I) \quad (\psi, \tau)$$

The fusion multiplicity matrices N for the product theories are just the product of the N matrices for the constituent theories

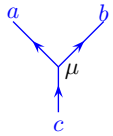
$$N^{(C,c)}_{(A,a),(B,b)} = N^C_{A,B}\, N^c_{a,b}$$

8.6 Tensor Description of Fusion and Splitting Spaces

Let us now try to give a bit more precise mathematical meaning to idea of fusion as well as to some of the diagrams we have been drawing[17]. For each fusion N^c_{ab} we define a space V^c_{ab} known as a fusion space and a space V^{ab}_c known as a splitting space. Both of these spaces have dimension N^c_{ab}

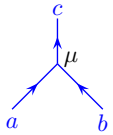
$$\dim V^c_{ab} = \dim V^{ab}_c = N^c_{ab}$$

Each of these spaces can be given an orthonormal basis, which we label with an index μ. We can write states in this space as kets which we draw as diagrams



$$= \quad |a, b; c, \mu\rangle \in V^{ab}_c \tag{8.16}$$

describes the splitting space. The Hermitian conjugate, the corresponding bras, are drawn as fusion diagrams



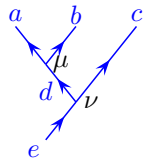
$$= \quad \langle a, b; c, \mu| \in V^c_{ab} \tag{8.17}$$

In the most commonly considered case, $N^c_{ab} = 1$ in which case there is a unique state and we do not need to specify μ since it has only one

[17] Those who feel they have a good understanding of the physics without needing the mathematics may be able to skip this section.

possible value. Cases where $N_{ab}^{c} = 0$ area non-allowed fusions meaning that the space V_{ab}^{c} and V_{c}^{ab} are zero dimensional.

In Eq. 8.16 we have described states in the space associated with a single anyon splitting into two, and in Eq. 8.17 we have described states in the space associated with two anyons fusing into one. It is possible to also describe the splitting or fusion space for a single anyon splitting or fusing into more pieces. For example, when splitting/fusing into three pieces the relevant spaces are often denoted[18] as V_{e}^{abc} or the Hermitian conjugate V_{abc}^{e}. A basis of states in this space can be written diagrammatically with the above described splitting vertices as[19]



$$= |(a,b);d,\mu\rangle \otimes |d,c;e,\nu\rangle \in V_{d}^{ab} \otimes V_{e}^{dc} \in V_{e}^{abc} \tag{8.18}$$

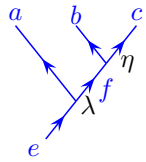
The full splitting space V_{e}^{abc} can thus be described as

$$V_{e}^{abc} \cong \bigoplus_{d} V_{d}^{ab} \otimes V_{e}^{dc} \tag{8.19}$$

with a corresponding dimension of this space

$$\dim V_{e}^{abc} = \sum_{d} N_{d}^{ab} N_{e}^{dc} \quad . \tag{8.20}$$

On the other hand, we could just as well have described a state in this space as



$$= |a,f;e,\lambda\rangle \otimes |(b,c);f,\eta\rangle \in V_{e}^{af} \otimes V_{f}^{bc} \in V_{e}^{abc} \tag{8.21}$$

In this language the full splitting space V_{e}^{abc} can be described as

$$V_{e}^{abc} \cong \bigoplus_{f} V_{e}^{af} \otimes V_{f}^{bc} \tag{8.22}$$

with a corresponding dimension of this space

$$\dim V_{e}^{abc} = \sum_{d} N_{e}^{af} N_{f}^{bc} \quad . \tag{8.23}$$

We thus have

$$V_{e}^{abc} \cong \bigoplus_{d} V_{d}^{ab} \otimes V_{e}^{dc} \cong \bigoplus_{f} V_{e}^{af} \otimes V_{f}^{bc} \tag{8.24}$$

where "≅" means "isomorphic to". In other words, these are two iso-

[18] We do not mean e to necessarily be the identity here. See note 1 of this chapter. We use this notation to match that of the next chapter.

[19] The insertion of the parenthesis (a,b) in Eq. 8.18, and similarly the parenthesis (b,c) in Eq. 8.21 are crucial to indicate which splitting is closest to the leaves of the tree (furthest from the root). Without the parenthesis one can have ambiguous notation, such as in the Fibonacci theory, where $|\tau\tau;\tau\rangle \otimes |\tau\tau;\tau\rangle$ could mean either the state in Eq. 8.18 or Eq. 8.21. The notation is telling us something important: that the kets in Eq. 8.18 and Eq. 8.21 are living in different, albeit isomorphic, spaces.

morphic decriptions of the same space. Equating the two different expressions (Eq. 8.20 and 8.23) for the dimension of this space recovers the equality Eq. 8.10. The isomorphism between these two descriptions of the same space will be explored in detail in the next chapter.

One can describe more complicated splitting and fusion spaces in an analogous way. For example, the space V_e^{aaaaa} can be described as

$$V_e^{aaaaa} \cong \bigoplus_{b,c,d} V_b^{aa} \otimes V_c^{ba} \otimes V_d^{ca} \otimes V_e^{da}$$

where each term in the direct sum (i.e., each term with fixed b, c, d) is drawn diagrammatically as in Fig. 8.14.

Chapter Summary

- This is

Further Reading

This is some reading.

Exercises

Exercise 8.1 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

$$\mathsf{d}_a \mathsf{d}_b = \sum_c N_{ab}^c \, \mathsf{d}_c$$

We will prove this formula algebraically in Chapter 17. However there is a simple and much more physical way to get to the the result: Imagine fusing together M anyons of type a and M anyons of type b where M gets very large and determine the dimension of space that results. Then imagine fusing together $a \times b$ and do this M times and then fuse together all the results.

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. 8.16, we must match the quantum number on one open end to the (opposite) quantum number on the opposite open 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. 8.18. 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.

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

(c) Show that in the limit of large number of handles M the ground state degeneracy scales as $\sim \mathcal{D}^M$ where $\mathcal{D}^2 = \sum_a \mathsf{d}_a^2$.

9

Change of Basis and F-Matrices[1]

[1]This chapter is crucial for the understanding of topological quantum systems. If there is one chapter to really study closely, this one is it! Don't worry too much about the section on gauge transforms or the appendices.

Let us consider the case of three anyons a, b and c that fuse together to form an anyon e. As mentioned several times previously (See Fig. 8.15 or Eqs. 8.18 and 8.21) one can describe the state of these three particles in two different ways. We can describe the space by describing how a fuses with b (the value of d on the left of Fig. 9.1), or by how b fuses with c (the value of f on the right of Fig. 9.1). Either of these two descriptions should be able to describe any state of the three anyons a, b and c fusing to e. However, in the two different cases these states are described in different bases. We define the change of basis as a set of unitary matrices[2,3,4] called F, as shown in Fig. 9.1.

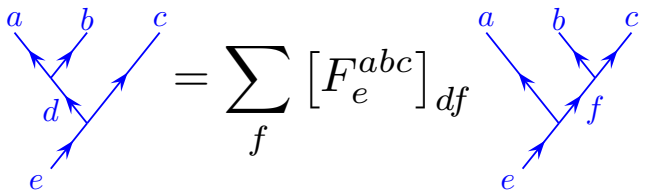


Fig. 9.1 The F-matrix makes a change of basis between the two different ways of describing the space spanned by the fusion of three anyon charges a, b, and c when they all fuse to a total quantum number of e. For fixed a, b, c and e, the matrix F is unitary in its subscripts d, f. Here F is defined to be zero if the fusion diagram is not allowed, i.e, if any of the fusion multiplicies N_{ab}^{d}, N_{dc}^{e}, N_{bc}^{f}, N_{af}^{e} are zero.

[2]For simplicity we are assuming no fusion multiplicities N_{ab}^{c} greater than 1. In cases where $N_{ab}^{c} > 1$ (as in Fig. 8.12), each vertex gets an additional index which ranges from 1 to its multiplicity so that the F matrix gets additional indices as well. This case is discussed in section 9.5.3.

[3]The conventions for writing F-matrices used in this chapter match that of Refs. Kitaev [2006] and Bonderson [2007].

Several brief comments are in order. First, as noted in the caption of Fig. 9.1 the F matrix is considered to be zero if any of the vertices on either side of the diagram are not allowed vertices of the fusion algebra. Secondly F moves involving the identity particle (i.e, with a, b or c being the identity in the figure) are chosen to have value of unity[5]. In particular this means

$$[F_e^{Ibc}]_{be} = [F_e^{aIc}]_{ac} = [F_e^{abI}]_{eb} = 1 \tag{9.1}$$

[5]This involves a gauge choice, see section 9.4.

[4]In the notation of section 8.6 the F matrix describes the isomorphism $V_d^{ab} \otimes V_e^{dc} \cong V_e^{af} \otimes V_f^{bc}$. We can write the basis change more algebraically as

$$|(a,b);d\rangle \otimes |d,c;e\rangle = \sum_f [F_e^{abc}]_{df} \ |a,f;e\rangle \otimes |(b,c);f\rangle$$

which represents Fig. 9.1 (where we have again suppressed indices $\mu, \ldots$ at the vertices for simplicity).

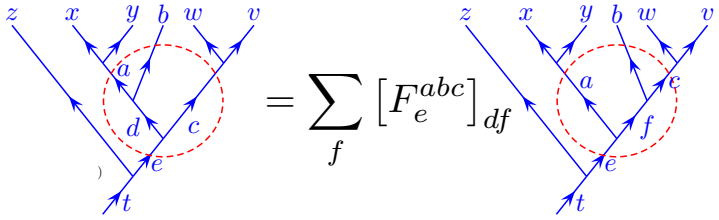


Fig. 9.2 The F-matrix can be applied inside of more complicated diagrams. Outside of the red circle both diagrams are the same. Inside the circle there is exactly the same transformation as is shown in Fig. 9.1.

Finally, being a change of basis, the F matrix (for fixed a, b, c, e) is unitary viewed as a matrix with indices d and 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 describe this space in terms of a combined with b to get d (as in the left of Fig. 9.1) or in terms of b combined with c to get f (as in the right of Fig. 9.1). In fact, even when studying TQFTs, sometimes people refer to F-matrices as $6j$ symbols.

It is important to emphasize that an F-matrix can act on a portion of a diagram, as shown in Fig. 9.2. This allows us to convert any treestructure in a fusion diagram to any other tree structure.

9.1 Example: Fibonacci Anyons

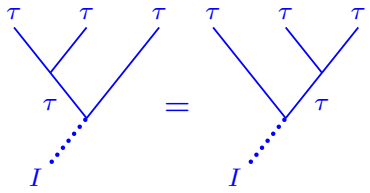
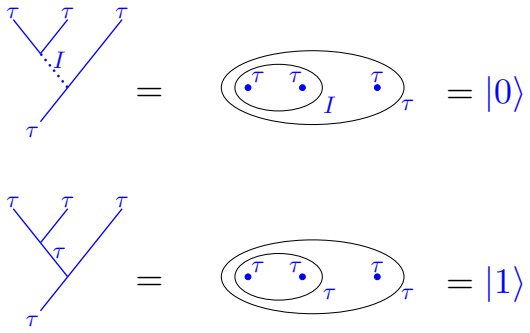


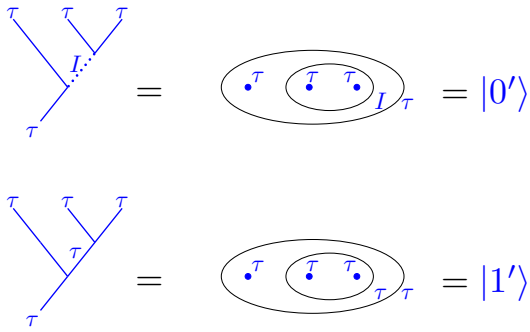
Fig. 9.3 There is only one state in the Hilbert space of three Fibonacci anyons fusing to the identity (we previously called this $|N\rangle$. Thus it does not matter if you fuse the left two first or the right two first, you are describing the same state.

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. 9.3. Mathematically we would write that $F_I^{\tau\tau\tau} = 1$. (And as noted in Eq. 9.1, if any of the three upper indices are the identity, we also have $F = 1$). The more interesting situation is the case where the three Fibonacci anyons fuse to τ. In this case, there is a two dimensional space of states, and this two dimensional space can be described in two ways. We can fuse the left two particles first to get either I (yielding overall state $|0\rangle$) or to get τ (yielding overall state $|1\rangle$). See the top of Fig. 9.4. 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. 9.4.

The space of states spanned by the three anyons is the same in either



Fusing the two particles on the left first



Fusing the two particles on the right first

Fig. 9.4 Two ways to describe the same two dimensional space in the case of Fibonacci anyons. 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.

description. Thus, there must be a unitary basis transform given by

$$\begin{pmatrix} |0\rangle \\ |1\rangle \end{pmatrix} = \begin{pmatrix} F_{00'} & F_{01'} \\ F_{10'} & F_{11'} \end{pmatrix} \begin{pmatrix} |0'\rangle \\ |1'\rangle \end{pmatrix} \tag{9.2}$$

Here F is a two by two matrix, and in the notation of the F matrix defined in Fig. 9.1, 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 we have used abbreviated notation here for more clarity.

For the Fibonacci theory the F matrix is given explicitly by[6]

$$F_\tau^{\tau\tau\tau} = F = \begin{pmatrix} \phi^{-1} & \phi^{-1/2} \\ \phi^{-1/2} & -\phi^{-1} \end{pmatrix} \tag{9.3}$$

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. In Section 9.3 we will discuss how this matrix is derived (See also section 18.2).

[6] We can redefine kets with different gauge choices (see section 9.4) and this will insert some phases into the off-diagonal of this matrix, but the simplest gauge choice gives the matrix as shown.

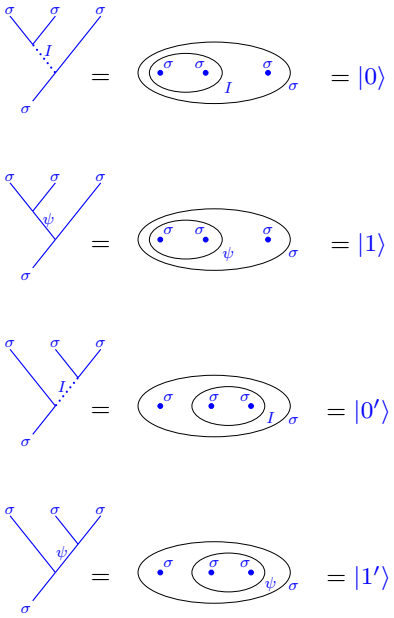


Fig. 9.5 Fusing three σ particles in the Ising theory. In $|0\rangle$ and $|1\rangle$ we fuse the left two particles first, whereas in $|0'\rangle$ and $|1'\rangle$ we fuse the right two particles first.

[7] It is interesting that the first of these two equations is a gauge independent statement, whereas the second of these two equations involves a gauge choice. See section 9.4 and exercise 9.1.

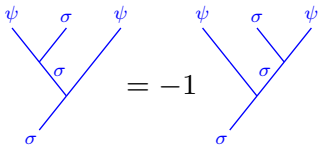


Fig. 9.6 Diagrammatic representation of Eq. 9.5. Both diagrams describe the same state in the Hilbert space, but they differ in a -1 phase.

9.2 Example: Ising Anyons

The situation with Ising anyons is quite similar, so we will be rather brief. Let us fuse three σ particles to an overall fusion channel of σ. There is no other choice, three σ particles can *only* fuse to σ (I.e., there is no $|N\rangle$ state. See section 8.2.2). There are two possible states in the Hilbert space which we can write in either of two bases as shown in Fig. 9.5 — either fusing the left two particles first or fusing the right two particles first. Analogous to the Fibonacci case we can write an F-matrix which relates the two basis descriptions as in Eq. 9.2. However, here the F matrix is instead given by

$$F_\sigma^{\sigma\sigma\sigma} = F = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{9.4}$$

which is sometimes known as a *Hadamard* matrix. Deriving this form of the F-matrix will be described roughly in section 9.3 below, and is done in detail in sections 18.3 and 19.4 below (See also exercise 9.7).

In the Ising theory we can also look at situations where we have both σ and ψ particles. In this case we have[7]

$$[F_\sigma^{\psi\sigma\psi}]_{\sigma\sigma} = -1 \tag{9.5}$$
$$[F_\psi^{\sigma\psi\sigma}]_{\sigma\sigma} = -1 \tag{9.6}$$

Eq. 9.5 is shown diagrammatically in Fig. 9.6. The other elements of F in the Ising theory which we have not mentioned so far (i.e., those not described by Eqs. 9.4-9.6) are either 1 if all the fusion vertices are allowed, or are zero if any of the fusion vertices are not allowed (See Fig. 9.1 caption).

The presence of the minus signs in Eqs. 9.5 and 9.6 may seem a bit puzzling being that, for example, the diagrams on the left and right of Fig. 9.6 are describing the same state in the Hilbert space. However, we will see in the next section why this sign is required in order to have a consistent F-matrix (See exercise 9.3 for a more detailed calculation).

9.3 Pentagon

It is possible to describe the same Hilbert space in many ways. For example, with three anyons, as in Fig. 8.15, 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. 9.1, 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. **??**. The fact that we can change the connectivity of these tree diagrams then allows one to make multiple changes in the trees as shown in Fig. 9.7 Indeed, in this figure one sees that one can go from the far left

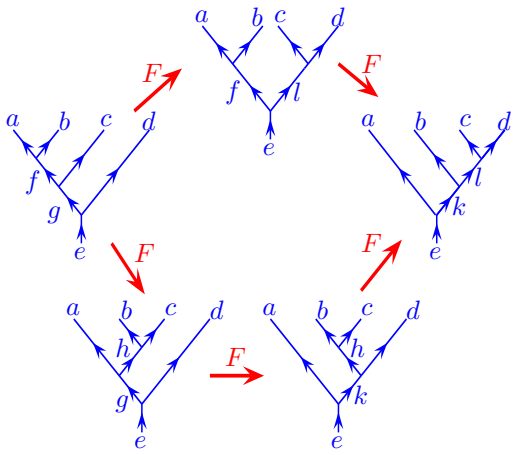


Fig. 9.7 Pentagon Diagram. Each step in the diagram is a new description of the same basis of states via and F-matrix.

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. This diagram, known as the pentagon diagram[8], puts a very strong contraint on the F matrices, which written out algebraically would be

$$[F_e^{fcd}]_{gl}[F_e^{abl}]_{fk} = \sum_h [F_g^{abc}]_{fh}[F_e^{ahd}]_{gk}[F_k^{bcd}]_{hl} \tag{9.7}$$

where the left hand side represents the top route of the figure and the right hand side represents the bottom route.[9]

For very simple theories, such as the Fibonacci anyon theory, the fusion rules and the Pentagon diagram are sufficient to completely define the F-matrices (up to some gauge convention choices as in section 9.4). See exercise 9.4. Further, for *any* given set of fusion rules there are a finite set of possible solutions of the pentagon equation[10] — a property that goes by the name "Ocneanu rigidity"[11].

One might think that one could write down more complicated trees and more complicated paths through the trees analogous to Fig. 9.7 and somehow derive additional constraints on the F-matrices. A theorem by MacLane [1971], known as the "coherence theorem", guarantees that no more complicated trees generate new identities beyond the pentagon diagram.

[8] An analogous relation holds for $6j$ symbols of angular momentum addition, known often as the Elliot-Biedenharn identity.

[9] It is very worth working through this to make sure you understand how this equation matches up with the figure! Note that in the equation the F matrices are written in an order such that those furthest right in Fig. 9.7 are furthest right in the equations.

[10] A finite set of *gauge inequivalent* solutions. I.e., a gauge transform of a given solution does not count as a new solution.

[11] Ocneanu did not manage to ever publish this important result. See for example Etingof et al. [2005].

9.4 Gauge Transforms

We have the freedom to make gauge transformations on our diagrams and these will be reflected in the F-matrix. While this is a bit of a technical point, we make frequent use gauge transformation in some later chapters so it is worth discussing it briefly here.

A gauge choice is a choice of a phase associated with the vertices in a diagram. If we chance this gauge choice, diagrams are then multiplied by phases. Let us define a second vertex

In particular a gauge transformation multiplies the vertices in a diagram by a phase as shown in Fig. 9.8. The tilde over the vertex on the right notates that we have made a gauge transform to a tilde gauge[12]

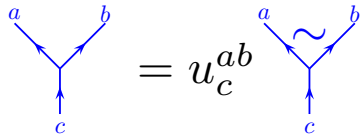


Fig. 9.8 We have the freedom to make a gauge transform of a vertex by multiplying by a phase u_c^{ab}. The tilde on the right notates that the vertex is in the tilde gauge.

Under such gauge transforms, the F-matrix must correspondingly transform as

$$\widetilde{[F_e^{abc}]}_{df} = \frac{u_e^{af} u_f^{bc}}{u_d^{ab} u_e^{dc}} [F_e^{abc}]_{df} \tag{9.8}$$

As we shall see in section 18.1.2, some gauge choices are much more natural than others, but we should always keep in mind that we have this freedom.

Note that if one of the upper legs is the identity ($a = I$ or $b = I$ in Fig. 9.8) we typically do not allow a gauge transform of this type of vertex, since the presence of a vertex with the vacuum is the same as the absence of a vertex with the vacuum (i.e, we can add or remove lines labeled by I for free).[13]

9.5 Appendix: F-matrix Odds and Ends

9.5.1 Product Theories

Given two anyon theories T and t, we can construct the product theory $T \times t$ as in section 8.5. If the theory T has consistent F matrices $[F_E^{ABC}]_{DF}$ and the theory t has consistent F matrices $[F_e^{abc}]_{df}$ ("consistent" here means satisfying the pentagon relation), then the product theory has consistent F matrices

$$[F^{(A,a)(B,b)(C,c)}_{(E,e)}]_{(D,d),(F,f)} = [F_E^{ABC}]_{DF} \; [F_e^{abc}]_{df}$$

The point here is that in a product theory, the two consituent theories don't "see" each other at all.

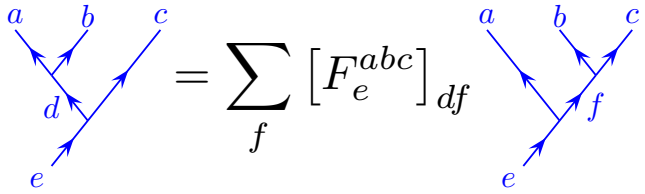
[12] This is much more easily expressed using the notation of section 8.6 where we can just write

$$|a, b; c\rangle = u_c^{ab} \widetilde{|a, b; c\rangle}$$

[13] There can be cases where we do want to specify that a vacuum line has branched off at one particular point and we do allow choosing a nontrivial gauge for such a vertex (See Lin and Levin [2014] for further discussion of this possibility).

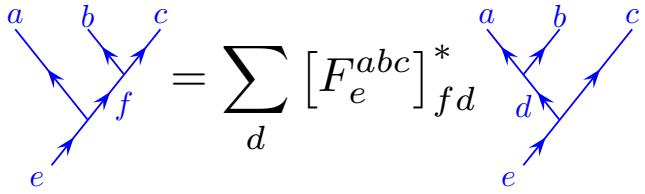
9.5.2 Unitarity of F

The F-matrix relation we defined as



$$= \sum_f \left[F_e^{abc}\right]_{df}$$

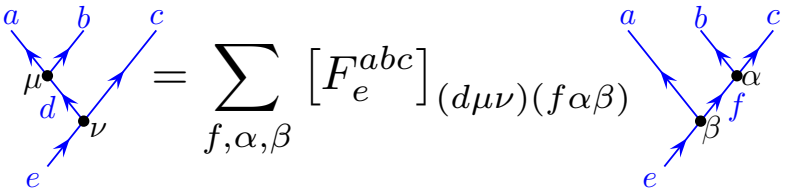
The fact that F is unitary in its indices d and f means we can also write



$$= \sum_d \left[F_e^{abc}\right]^*_{fd}$$

9.5.3 F-matrix with higher fusion multiplicities

In cases where there are fusion multiplicities N_{ab}^c greater than 1, each vertex gets an additional index as shown in Fig.8.12. The F-matrix must also describe what happens to these indices under basis transform. We thus have a more general basis-change equation given in Fig. 9.9.



$$= \sum_{f,\alpha,\beta} \left[F_e^{abc}\right]_{(d\mu\nu)(f\alpha\beta)}$$

Fig. 9.9 The F-matrix equation with fusion multiplicities greater than one. Here the vertex indices are $\mu \in 1 \ldots N_{ab}^d$ and $\nu \in 1 \ldots N_{dc}^e$ and $\alpha \in 1 \ldots N_{bc}^f$ and $\beta \in 1 \ldots N_{af}^e$. The subscripts $(d\mu\nu)$ and $(f\alpha\beta)$ are "super-indices", of the matrix F_e^{abc}. I.e., d, μ and ν are joined together to make a single index.

In the language of section 8.6 this F-transform is written as

$$|a,b;d,\mu\rangle \otimes |d,c;e,\nu\rangle = \sum_{f,\alpha,\beta} \left[F_e^{abc}\right]_{(d\mu\nu)(f\alpha\beta)} |a,f;e,\beta\rangle \otimes |b,c;f,\alpha\rangle$$

Gauge Transforms with higher fusion multiplicities

With higher fusion multiplicities $N_{ab}^c > 1$, our diagrams have indices at the vertices. Gauge transforms are generally a unitary matrix within this index space and then take the form shown in Fig. 9.10[14]

Under such gauge transforms, the F-matrix must correspondingly

[14] Again, this is much more easily expressed using the notation of section 8.6 where we can just write

$$|a,b;c,\mu\rangle = \sum_{\mu'} [u_c^{ab}]_{\mu\mu'} |\widetilde{a,b;c,\mu}\rangle$$

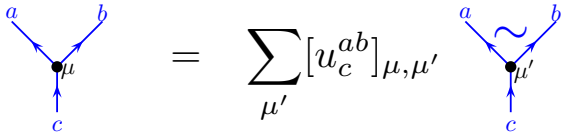


Fig. 9.10 We have the freedom to make a gauge transform of a vertex by multiplying by a unitary matrix $[u_c^{ab}]_{\mu\mu'}$. The tilde on the right notates that the vertex is in the tilde gauge.

transform as

$$\widetilde{[F_e^{abc}]}_{(d\mu'\nu')(f\alpha'\beta')} = \sum_{\alpha,\beta,\mu,\nu} ([u_d^{ab}]^{-1})_{\mu'\mu}([u_e^{dc}]^{-1})_{\nu'\nu}[F_e^{abc}]_{(d\mu\nu)(f\alpha\beta)}[u_e^{af}]_{\beta\beta'}[u_f^{bc}]_{\alpha\alpha'} \tag{9.9}$$

Chapter Summary

- This is the summary

Further Reading

Exercises

Exercise 9.1 *F*-**gauge choice**

(a) Explain why in the Fibonacci theory, $[F_\tau^{\tau\tau\tau}]_{\tau\tau}$ is gauge independent but $[F_\tau^{\tau\tau\tau}]_{I\tau}$ is gauge dependent.

(b) Explain why in the Ising theory is $[F_\sigma^{\psi\sigma\psi}]_{\sigma\sigma}$ is gauge independent, but $[F_\psi^{\sigma\psi\sigma}]_{\sigma\sigma}$ is gauge dependent.

Exercise 9.2 *F*'**s with the vacuum field** *I*

Explain why $[F_e^{aIc}]_{ac} = [F_d^{abI}]_{db} = [F_e^{Ibc}]_{be} = 1$.

Exercise 9.3 Ising Pentagon

Consider a system of Ising anyons. Given the fusion rules, F_w^{xyz} will be a 2 by 2 matrix in the case of $x = y = z = w = \sigma$ (given by Eq. 9.4) and is a simply a scalar otherwise. One might hope that these scalars can all be taken to be unity. Unfortunately this is not the case. By examining the pentagon equation, Eq. 9.7 in the case of $a = b = c = \sigma$ and $d = f = \psi$ show that taking the scalar to always be unity is not consistent. Show further that choosing $[F_\sigma^{\psi\sigma\psi}]_{\sigma\sigma} = -1$ (and leaving the other scalars to be unity) allows a consistent solution of the pentagon for $a = b = c = \sigma$ and $d = f = \psi$.

Exercise 9.4 Fibonacci Pentagon

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 up to one sign (gauge) choice. Using the pentagon equation determine the F-matrix. (To get you started, note that in Fig. 9.7 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 Nayak et al. [2008].

Exercise 9.5 Pentagon and Fusion Multiplicities

Consider the case of Appendix 9.5.3 where there are fusion multiplicities $N_{ab}^c > 1$. Write the generalization of the pentagon equation Eq. 9.7.

Exercise 9.6 Gauge Change

(a.i) Confirm that the F-matrix transforms under gauge change as indicated in Eq. 14.2. (a.ii) Show that a solution of the pentagon equation remains a solution under any gauge transformation.

[Harder] Now consider the case of Appendix 9.5.3 where there are fusion multiplicities $N_{ab}^c > 1$

(b.i) Analogous to (a.i) Confirm Eq. 9.9. (b.ii) Analogous to (a.ii) show that a solution of the pentagon equation remains a solution under any gauge transformation. (You will need to solve problem 9.5 first!)

Exercise 9.7 Ising F-matrix

[Hard] As discussed in the earlier problem, "Ising Anyons and Majorana Fermions" (Ex, 3.3), one can express Ising anyons in terms of Majorana fermions which are operators γ_i with anticommutations $\{\gamma_i, \gamma_j\} = 2\delta_{ij}$. 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.

10 Exchanging Identical Particles

We would now like to deterime what happens when two particles are exchanged with each other. As one might expect for anyons, phases are accumulated from such exchanges. However, one must be cautious because the phase accumulated will generally depend on the fusion channel of the particles being exchanged.

10.1 Introducing the R-matrix

Let us begin with a simple case where two identical particles of type a are braided around each other. Let us specify that the two particles fuse together in an overall channel c. Let us call this quantum mechanical state $|\text{state}\rangle$ as shown in two different notations in Fig. 10.1.

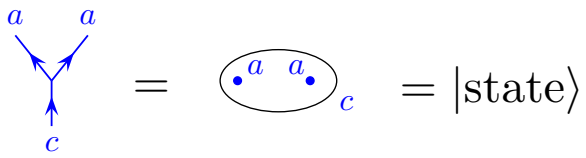


Fig. 10.1 Two a particles fusing to a c particle.

We then (half)-braid the two particles around each other (counter-clockwise observing from above[1].). The final fusion channel of the two a particles is still c (by the locality principle of section 8.2). However, a phase will be accumulated in the process which call R_c^{aa} as shown in Fig. 10.2. The inverse phase would be accumulated for an exchange in the opposite direction.

[1] In the language of the braid group we would call this exchange σ. See section 3.3.1.

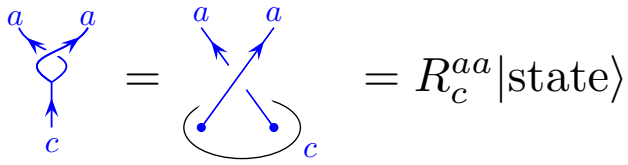


Fig. 10.2 The phase accumulated by exchanging two a particles that fuse to c is called R_c^{aa}.

This so-called R-matrix along with the corresponding F-matrices will allow us to compute the result of braiding any number of a particles around each other in arbitrary ways.

Let us consider the case of three anyons of type a. We can write a basis for the possible states of three anyons as shown in Fig. 10.3.

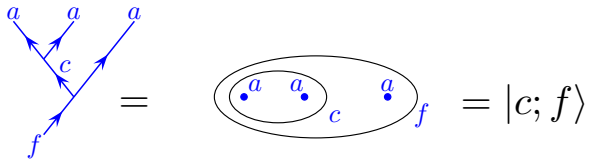


Fig. 10.3 A basis of states for three a type anyons fusing to an overall quantum number f.

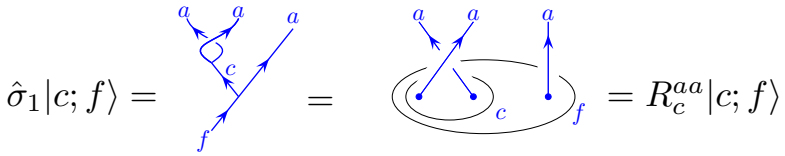


Fig. 10.4 Exchanging the two left particles incurs a phase R_c^{aa}.

We now consider exchanging the left two particles as shown in Fig. 10.4. (We call the operator that performs this exchange $\hat{\sigma}_1$ in analogy with the braid group discussed in section 3.3.1.) Since we know the fusion channel of these two particles (c) we know that the phase accumulated in this exchange is just R_c^{aa}. This seems fairly simple as it is precisely the type of exchange we defined in Fig. 10.2 above

As with all operators in quantum mechanics that can be implemented as a time evolution, the exchnge operator is linear, meaning that it acts on superpositions by acting on each term individually:

$$\hat{\sigma}_1 \sum_c \alpha_c |c; f\rangle = \sum_c \alpha_c R_c^{aa} |c; f\rangle$$

Let us now instead consider exchanging the right two particles, an operation we call $\hat{\sigma}_2$. Since the right two particles are not in a definite fusion channel we cannot directly apply the R-matrix. However, we can use the F-matrix to rewrite our state as a superposition of states where the right two particles are in a definite fusion channel as shown in Fig. 10.5.

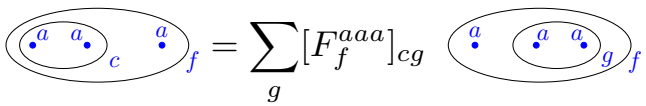


Fig. 10.5 Using an F-move to work in the basis with a known fusion channel of the right two particles.

One we have transformed to this new basis, then we can exchange the right two particles and apply the R-matrix directly to the right two particles as shown in Fig. 10.6. Once we have established the effect of the exchange we can (if desired) convert back into the original basis which describes the fusion of the left two particles using F^{-1}.

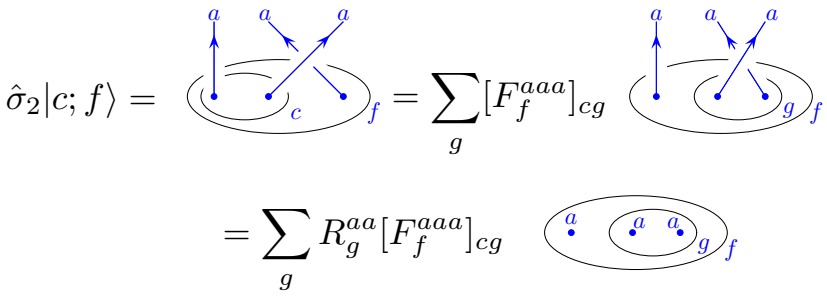


Fig. 10.6 In order to describe exchange of the right two particles, we first change to a basis where the fusion channel of those two particles is explicit. We can then apply the R matrix directly.

The result of this procedure in terms of the original basis, is given by

$$\hat{\sigma}_2|c;f\rangle = \sum_{g,z} [F_f^{aaa}]_{cg} \; R_g^{aa} \; [(F_f^{aaa})^{-1}]_{gz} \; |z;f\rangle \tag{10.1}$$

The general principle is that to evaluate any exchange of identical particles, we can always use F matrices to convert to a basis where the fusion channel of the two particles to be braided is known. Once we are working in this basis, we can then we apply the R matrix directly. At the end we can transform back to the original basis if we so desire. This scheme works for *any* set of identical particles given appropriate F and R matrices.

10.1.1 Locality

An important principle which we will often use is that result of braiding a group of particles with a given total quantum number c is the same as if that entire group were replaced with just a single particle with quantum number c. For example, in Fig. 10.7 when we braid a cluster of a, b with overall quantum number d around a cluster x, y, z with overall quantum number w, the phase accumulated should be the same as if we simply braided d around w.

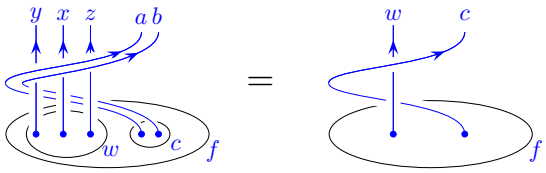


Fig. 10.7 Braiding a cluster of particles with overall quantum number c around a cluster of particles with overall quantum number w should have the same effect as braiding a single c particle, around a single w particle. The result should just be a phase dependent on the quantum numbers c, w and f. In chapter 13 we will refer to this phase as $R_f^{wc} R_f^{cw}$. Note that a does not wrap around b. If it did that would accumulate an additional phase.

10.2 Some Examples

Since the idea of using the R-matrix is quite important, it is worth working through a few examples explicitly.

10.2.1 Fibonacci Anyons

Recall the properties of Fibonacci anyons (see section 8.2.1): There is only one nontrivial particle type which we call τ and the only nontrivial fusion rule is $\tau \times \tau = I + \tau$. As we saw in section 8.2.1, the fusion rule imples that there are two possible states of two Fibonacci anyons: The state where they fuse together to form I and the state where they fuse together to τ (See Fig 10.8. See also Fig. 8.9 where we previsously introduced these two states). We call these states $|I\rangle$ and $|\tau\rangle$ respectively.

Now consider the operator $\hat{\sigma}$ that exchanges the two Fibonacci anyons counterclockwise as viewed from above, as shown in Fig. 10.9 This operator yields the phase $R_I^{\tau\tau}$ 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.

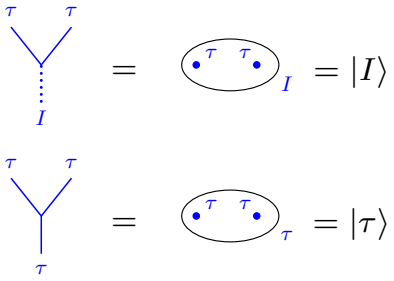


Fig. 10.8 The two possible states of two Fibonacci anyons. Note that we do not draw arrows on the particle lines in the left diagrams since τ is self-dual.

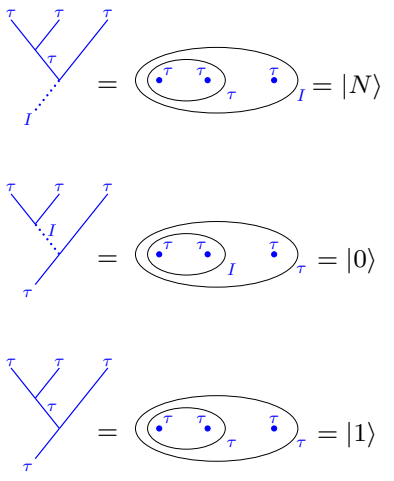


Fig. 10.10 The three states in the Hilbert space of three Fibonacci anyons.

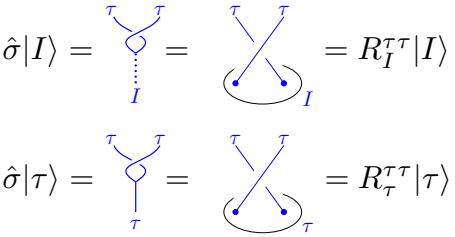


Fig. 10.9 Exchanging two anyons gives a phase dependent on their fusion channel.

In section 13.3 below (see also exercise 10.6) we will explain how we actually compute the phases $R_\tau^{\tau\tau}$ and $R_I^{\tau\tau}$. For now, it suffices to give the answers that for *right-handed* Fibonacci anyons

$$\begin{aligned} R_\tau^{\tau\tau} &= e^{+3\pi i/5} \\ R_I^{\tau\tau} &= e^{-4\pi i/5} \end{aligned} \tag{10.2}$$

There also exists a left-handed type of Fibonacci anyons for which the phases are complex conjugate of these.

As with all operators in quantum mechanics we can act on superpositions by acting on each term individually:

$$\hat{\sigma}\left(\alpha|I\rangle + \beta|\tau\rangle\right) = \alpha R_I^{\tau\tau}|I\rangle + \beta R_\tau^{\tau\tau}|\tau\rangle \quad .$$

If we think of our two states $|I\rangle$ and $|\tau\rangle$ as the two states of a qubit, the $\hat{\sigma}$ operator is what is known as a controlled phase gate in quantum information processing — the phase accumulated depends on the state of the qubit.

Next let us consider the possible states of three Fibonacci anyon. As

described in section 8.2.1 the space of such states is three-dimensional, and we can choose as a basis the two states shown in Fig. 10.10 (we already introduced these states in Fig. 8.10 above). Now consider an operator $\hat{\sigma}_1$ that braids the two leftmost particles around each other as shown in Fig. 10.11. Here the phase accumulated depends on the fusion channel of the leftmost two particles, entirely analogous to Fig. 10.9.

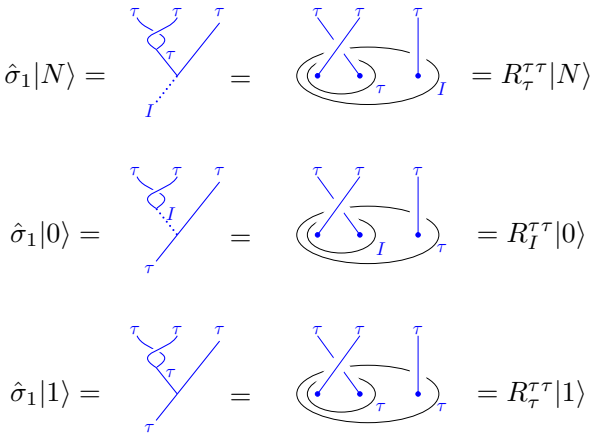


Fig. 10.11 Exchanging the left two particle.

More interesting is the question of what happens if we exchange the right two particles as shown in Fig. 10.12. As discussed in Section 10.1, 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.

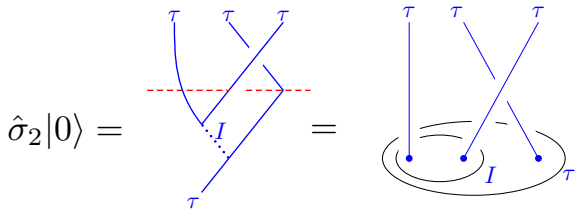


Fig. 10.12 Exchanging the two particles on the right for the $|0\rangle$ state where these two particles on the right are not in a definite fusion channel. Note that in the tree diagram on the left the state below the dashed red line is exactly $|0\rangle$.

To see how this works, Recall that we can use the F-matrix to write the $|0\rangle$ state in the basis of the $|0'\rangle$ and $|1'\rangle$ as in Eq. 9.2 which we reproduce the relevant parts of here:

$$|0\rangle = F_{00'}|0'\rangle + F_{01'}|1'\rangle \tag{10.3}$$

where $|0'\rangle$ and $|1'\rangle$ are shown in Fig. 10.13. Note that here F_{ab} is shorthand for $[F_\tau^{\tau\tau\tau}]_{ab}$.

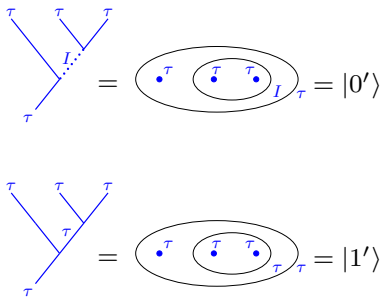
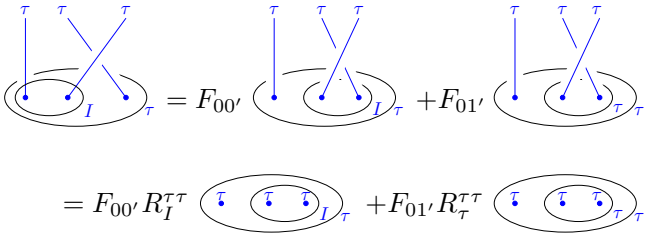


Fig. 10.13 In the prime basis the two particles on the right are in a definite fusion channel

On the right hand side of Fig. 10.13 (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 compute the corresponding phase as shown in Fig. 10.14.



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

Fig. 10.14 To exchange the right two particles we first use an F-move so that we know the fusion channel of these two particles, then we can apply R and then F^{-1} to transform back into the original basis.

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.[2]. The final result, Eq. 10.6 is precisely the same as Eq. 10.1 just written out in all of its detail[3]

We can summarize the results of the two possible braiding operations on the three dimensional Hilbert space. Assuming right-handed Fibonacci anyons and using a basis $|N\rangle, |0\rangle, |1\rangle$ (also notated as $|\tau; I\rangle$, $|I; \tau\rangle$, $|\tau; \tau\rangle$) we have

$$
\hat{\sigma}_1 = \begin{pmatrix} e^{3\pi i/5} & & \\ & e^{-4\pi i/5} & \\ & & e^{3\pi i/5} \end{pmatrix} \qquad (10.6)
$$

$$
\hat{\sigma}_2 = \begin{pmatrix} e^{3\pi i/5} & & \\ & \phi^{-1} e^{4\pi i/5} & \phi^{-1/2} e^{-3\pi i/5} \\ & \phi^{-1/2} e^{-3\pi i/5} & -\phi^{-1} \end{pmatrix} \qquad (10.7)
$$

where $\phi = (\sqrt{5}+1)/2$ is the golden mean.

10.2.2 Ising Anyons

For Ising anyons the situation is perhaps even simpler since three σ particles have only two fusion channels (See section 8.2.2). The appropriate F-matrices are given by Eq. 9.4 and the R-matrices for a *right-handed*

[2] For this particular case (using Eq. 9.3 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!)

[3] To fully harmonize the notation with that of Eq. 10.1 we should make the identification $|0\rangle \to |I;\tau\rangle$ and $|1\rangle \to |\tau;\tau\rangle$. The indices 0 and 1 are replaced by I and τ and as mentioned above the F_{ab} matrix is really $[F_\tau^{\tau\tau\tau}]_{ab}$.

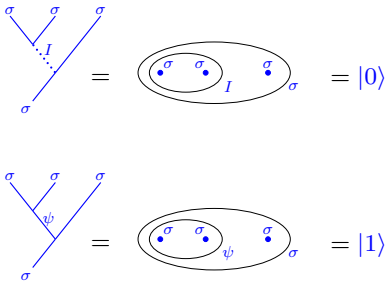


Fig. 10.15 A simple basis for a qubit made from three Ising anyons. (See Fig. 9.5).

Ising theory are given by

$$R_I^{\sigma\sigma} = e^{-i\pi/8} \tag{10.8}$$
$$R_\psi^{\sigma\sigma} = e^{i3\pi/8} = iR_I^{\sigma\sigma} \tag{10.9}$$

with the R-matrices for a left-handed theory being the complex conjugates of these expressions. From the R-matrix, we immediately obtain the form of the exchange operator $\hat{\sigma}_1$, that counterclockwise exchanges the leftmost two Ising anyons

$$\hat{\sigma}_1 = e^{-i\pi/8} \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}. \tag{10.10}$$

Then using Eq. 10.1 we can evaluate the exchange operator $\hat{\sigma}_2$ which counterclockwise exchanges the rightmost two anyons of the three, giving

$$\hat{\sigma}_2 = \frac{e^{i\pi/8}}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ -i & 1 \end{pmatrix}. \tag{10.11}$$

This is some reading

Exercises

Exercise 10.1 Calculating Exchanges

(a) Use Eq. 10.1 to confirm Eq. 10.11

(b) Use Eq. 10.1 to confirm Eq. 10.7

(c) Confirm the braiding relation $\hat{\sigma}_1\hat{\sigma}_2\hat{\sigma}_1 = \hat{\sigma}_2\hat{\sigma}_1\hat{\sigma}_2$ in both cases. What does this identity mean geometrically. See exercise 3.1.

Exercise 10.2 Ising Anyons Redux

In exercise 3.3 we introduced a representation for the exchange matrices for Ising anyons which, for three anyons, would be of the form

$$\hat{\sigma}_1 = \frac{e^{i\alpha}}{\sqrt{2}}(1 + \gamma_1\gamma_2) \tag{10.12}$$
$$\hat{\sigma}_2 = \frac{e^{i\alpha}}{\sqrt{2}}(1 + \gamma_2\gamma_3) \tag{10.13}$$

where the γ's are Majorana operators defined by

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

with $\gamma_i = \gamma_i^\dagger$.

Show that the exchange matrices in Eq. 10.11 are equivalent to this representation. How does one represent the $|0\rangle$ and $|1\rangle$ state of the Hilbert space

in this language? The answer may not be unique.

Exercise 10.3 Exchanging More Particles

(a) Consider a system of 4 identical Ising anyons. Use the F and R-matrices to calculate the braid matrices $\hat{\sigma}_1, \hat{\sigma}_2$, and $\hat{\sigma}_3$. (You should be able to check your answer using the Majorana representation of exercise 3.3.)

(b) [Harder] Consider a system of 4 identical Fibonacci anyons. Use the F and R-matrices to calculate the braid matrices $\hat{\sigma}_1, \hat{\sigma}_2$, and $\hat{\sigma}_3$.

Exercise 10.4 Determinant and Trace of Braid Matrices

Consider a system of N-identical anyons with a total Hilbert space dimension D. The braid matrix $\hat{\sigma}_1, \hat{\sigma}_2, \ldots, \hat{\sigma}_{N-1}$ are all D-dimensional. Show that each of these matrices has the same determinant, and each of these matrices has the same trace. Hint: This is easy if you think about it right!

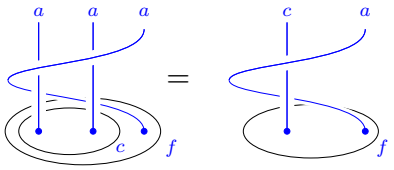


Fig. 10.16 The locality constraint (see similar figure 10.7).

Exercise 10.5 Checking the locality constraint

[Easy] Consider Fig. 10.16. The braid on the left can be written as $\hat{b}_3 = \hat{\sigma}_2\hat{\sigma}_1^2\hat{\sigma}_2$.

(a) For the Fibonacci theory with $a = \tau$ check that the matrix $\hat{b}_3$ gives just a phase, which is dependent on the fusion channel c. I.e., show the matrix $\hat{b}_3$ is a diagonal matrix of complex phases. Show further that these phases are the same as the phase that would be accumulated for taking a single τ particle around the particle c.

(b) Consider the same braid for the Ising theory with $a = \sigma$. Show again that the result is a c-dependent phase.

[Hard] Consider the braid shown on the left of Fig. 10.17. The braid can be written as $\hat{b}_4 = \hat{\sigma}_3\hat{\sigma}_2\hat{\sigma}_1^2\hat{\sigma}_2\hat{\sigma}_3$.

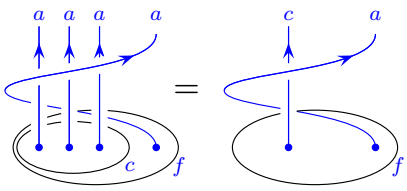


Fig. 10.17 The locality constraint (See similar figure 10.7).

(c) Consider Ising anyons where $a = \sigma$, Use the F and R-matrices to calculate $\hat{\sigma}_3$ (See exercise 10.3.a). Since the fusion of three σ anyons always gives $c = \sigma$, calculate $\hat{b}_4$, show this is a phase times the identity matrix, and show that the phase matches the phase of taking a single σ all the way around another σ.

(d) Consider Fibonacci anyons with $a = \tau$, Use the F and R-matrices to calculate $\hat{\sigma}_3$. (See exercise 10.3.b). Check that $\hat{b}_4$ is a diagonal matrix of phases. Check the phases match the two possible phases accumulated by wrapping a single τ all the way around a single particle c which can be I or τ.

Exercise 10.6 Enforcing the locality constraint

The locality constraint shown in Fig. 10.16 turns out to be extremely powerful. In this exercise we will use this constraint to (almost) derive the possible values for the R-matrix for Fibonacci anyons given the known F-matrix.

Consider an anyon theory with Fibonacci fusion rules and Fibonacci F-matrix as in Eq. 9.2.

(a) [Easy] Confirm the locality constraint shown in Fig. 10.16 (see also Fig. 10.7) given the values of R given in Eq. 10.2. Make sure to confirm the equality for all three cases $f = I, c = \tau$ and $f = \tau, c = I$ and $f = \tau, c = \tau$.

Note that on the left of Fig. 10.16 is the braiding operation $\hat{O} = \hat{\sigma}_2\hat{\sigma}_1\hat{\sigma}_1\hat{\sigma}_2$. whereas the operation on the right is σ^2.

(b) Show that the locality constraint of Fig. 10.16 would also be satisfied by

$$R_I^{\tau\tau} \to -R_I^{\tau\tau} \qquad R_\tau^{\tau\tau} \to -R_\tau^{\tau\tau} \tag{10.14}$$

It will turn out (See *** below) that this additional solution is spurious, as there are other consistency conditions it does not satisfy.

(c) In addition to right and left handed Fibonacci anyons and the two additional spurious solutions provided by Eq. 10.14, there are four additonal possible sets of R-matrices that are consistent with the F-matrices of the Fibonacci theory given the locality constraint of Fig. 10.16. These additional solutions are all fairly trivial. Can you guess any of them?

If we cannot guess the additional possible R-matrices, we can derive them explicitly (and show that no others exist). Let us suppose that we do not know the values of the R-matrix elements $R_I^{\tau\tau}$ and $R_\tau^{\tau\tau}$.

(d) For the case of $f = I$ and $c = \tau$ show that Fig. 10.16 implies

$$[R_\tau^{\tau\tau}]^4 = [R_I^{\tau\tau}]^2 \tag{10.15}$$

(e) [Harder] For the case of $f = \tau$ we have a two-dimensional Hilbert space spanned by the two values of $c = I$ or $c = \tau$. Any linear operator on this Hilbert space should be a 2 by 2 matrix. Thus the localtiy constraint Eq. 10.16 is actually an equality of 2 by 2 matrices. Derive this equality.

(f) Use this result, in combination with Eq. 10.15 to find all possible R-matrices that satisfy the localtiy constraint. You should find a total of eight solutions. Six of these are spurious as we will see in section ***.

The calculation you have just done is equivalent to enforcing the so-called hexagon condition which we will discuss in section *** below.

Computing with Anyons

11

Having discussed the basics of anyon theories, we are now in a position to discuss how one might perform quantum computations with braids.

In chapter 2 we briefly introduced some ideas of topological quantum computation. In chapter 8 we discussed how we might define a qubit in several simple anyon theories. In the current chapter we will briefly discuss how anyons can be used to fulfill the requirements for quantum computation[1].

[1] For more of the basics of quantum computation, a classic reference is Nielsen and Chuang [2000]. We also provide a bit more information in appendix chapter ***.

11.1 Quantum Computing

To have a quantum computer, we must first have a Hilbert space, and we usually think of this Hilbert space as being built from small pieces, such as qubits or qutrits[2]. This Hilbert space will be the quantum memory that the computer acts on.

Once we have our Hilbert space, our model of a quantum computer has three key steps for quantum computation[3]:

(0) Find a Hilbert space to work with.

(1) Initialize the Hilbert space in some known state.

(2) Perform a controlled unitary operation on the Hilbert space.

(3) Measure some degree of freedom in the Hilbert space.

If the controlled unitary (step 2) is implemented as a series of unitary operations each of which acts on only small parts of the Hilbert space (such as acting on just a few qubits at a time), we call this scheme for quantum computation the *quantum circuit model.*[4]

We will discuss each of the above steps (0)-(3) for our anyon systems in section 11.2 below. First, however, we will introduce the idea of what it means for a quantum computer to be "universal" in the quantum circuit model.

[2] Qubits are two state systems (such as a spin-$\frac{1}{2}$), qutrits are three state systems etc. The general case is known as a qudit. See the introduction to quantum information in appendix chapter ***

[3] There are variants on this theme. For example, it might be sufficient to initialize into a state that is only partially known, or it might be sufficient to have a somewhat noisy measurement. Most interesting is the issue of whether one can tolerate some amount of imperfection in the system (noise in the system, uncontrolled operations on the Hilbert space, etc). We will discuss this issue further in section **

[4] There are other models of quantum computation. We mention in particular the measurement schemes (See Raussendorf and Briegel [2001]; Gross et al. [2007]), where no unitary is explicitly performed, but rather the computation is implemented as a series of measurements on an initial highly entangled state. In the context of topological quantum computation an important variant is a computation that is implemented by a combination of unitary operations and projective measurements. The earliest proposal for quantum computing with anyons, by Kitaev in 1997, was of this type (See Kitaev [2003]). See also footnote 15 below.

11.1.1 Universal Quantum Computing in the Quantum Circuit Model

Let us suppose our Hilbert space consists of N qubits (each qubit being a two state system). The Hilbert space dimension is then $D = 2^N$. The space of possible unitary[5] operations[6] on these qubits, is just the group of D dimensional unitary matrices — a group known as $U(D)$.

Let us now suppose our quantum computer can implement any one of p different elementary operations (usually called "gates") in a single time step (each gates will act only on a small number of qubits). Each gate corresponds to a particular unitary operation $U_n \in U(D)$ with $n \in 1, \ldots, p$ that is applied to the Hilbert space. A sequence of such gates constructs a particular unitary operation which is just the product of the successive gates (the time order runs from left to right)

$$U = U_{i_t} \ldots U_{i_2} U_{i_1} \tag{11.1}$$

where the number of gates t, can be thought of as the "run time" of the computation.

Suppose there is some particular computation we would like to perform, and this computation corresponds to a unitary U which we hope to construct via a series of gates as in Eq. 11.1. Note, however, that in quantum computation we are never worried about the overall phase of our result.[8] As such if we want to construct some particular unitary U, for the purpose of quantum computation, it just as good to construct $e^{i\phi}U$ for any value of ϕ.

Unfortunately, even with this freedom of phase most unitary operations (except for a set of measure zero) are actually impossible to construct exactly from a finite set of elementary gates as in Eq. 11.1. Fortunately, for computational purposes it is good enough to *approximate* the desired unitary operation to some (potentially high) accuracy. Sets of gates that can always make such an accurate approximation are called *universal*. We will be more precise about the definition of this word in a moment.

Since we will be discussing approximations of desired operations, it is useful to define a distance between two unitary matrices so we can measure the accuracy of our approximation. Given two D dimensional unitary matrices U and V, we define a phase invariant distance measure between them as[9]

$$\mathbf{dist}(U; V) = \sqrt{1 - \frac{|\mathrm{Tr}[U^\dagger V]|}{D}} \tag{11.2}$$

Note that multiplying either matrix by an overall phase leaves **dist** unchanged, and if U and V are the same up to a phase, then **dist** is zero. We say that V is a good approximation of U up to a phase if $\mathbf{dist}(U; V)$ is small .

Having defined this distance measure, we can be more precise about what we mean that a set of gates is universal. A gate set $U_n \in U(D)$

[5] Recall a matrix U is unitary if and only if $U^\dagger U = U U^\dagger = \mathbf{1}$.

[6] Quantum mechanical time evolution is always unitary. This is simply the statement that a normalized ket remains normalized. It is worth noting that we are excluding the possibility of making measurements (which are generally nonunitary[7]) on the system before the end of the computation. This would be outside of the quantum circuit model.

[7] All of quantum mechanics can be viewed as unitary time evolution. Measurements may look like they are nonunitary, but one can always include the measuring apparatus within the system being considered and then the full system (including the measuring apparatus) then obeys unitary evolution. The idea of including measurement within your system in order to maintain unitarity is sometimes known as "the church of the larger Hilbert space".

[8] Recall that the last step of a quantum computation, after applying a unitary U (via some sequence of gates as in Eq. 11.1) to our Hilbert space, we obtain an output "answer" by measuring whether some particular qubits are in the $|0\rangle$ state or the $|1\rangle$ state. The probabilities of theses outcomes is completely independent of the overall phase of the U. I.e., if we changed $U \to e^{i\phi}U$ we would have the same probabilities of outcomes.

[9] I believe this distance measure was introduced by Fowler [2011]. Other definitions of distance can also be used. Relationships between this distance measure and more conventional operator norms are given by Field and Simula [2018] and Amy [2013].

with $n \in 1, \ldots, p$ is universal if for any any desired target operation we would like to perform $U_{\text{target}} \in U(D)$ we can find a sequence of gates $U_{i_1} U_{i_2} \ldots U_{i_t}$ such that the phase invariant distance to the target is less than any desired error tolerance ϵ

$$\mathbf{dist}(U_{\text{target}} \; ; \; U_{i_1} U_{i_2} \ldots U_{i_t}) < \epsilon \tag{11.3}$$

no matter how small an ϵ we choose. In other words, our gate set can approximate any target unitary as precisely as we want.

We might wonder how long a run time (how many gates) will we typically need to have? A beautiful theorem by Kitaev and Solovay[10] assures us that the run time is not too long[11]. In particular,

$$t \sim \mathcal{O}\left(\log(1/\epsilon)\right) \tag{11.4}$$

We are thus guaranteed that if we have a universal gate set, then the run time of the computer gets at most logarithmically longer as we try to increase the quality of our approximation of the target operation U_{target}.

The essence of this theorem is as follows. If we consider a sequence of t gates, (i.e., a run time of t), if there are p different elementary gates, we can construct roughly p^t different sequences of gates[12]. Thus as t gets larger, there are exponentially more possible unitaries we can construct and these roughly cover the space $U(D)$ evenly. With the number of points we can construct in this space growing exponentially with t, the distance ϵ of an arbitrary target unitary to the nearest unitary we can construct must drop exponentially with t, hence justifying Eq. 11.4.

It is a nontrivial calculation to determine which set of elementary gates is sufficient to have a quantum computer which is universal. However, an important result is that if one can perform arbitrary rotations on a single qubit and in addition if one can perform *any* entangling two-qubit operation between any of these two bits (or even between just neighboring bits), then one has a universal quantum computer[13].

[10] The Kitaev-Solovay theorem, often viewed as one of the most fundamental results of quantum computation, is discussed nicely in Dawson and Nielsen [2006] and Harrow [2001].

[11] The usual proof of the Kitaev-Solovay theorem assumes that the gate set must contain inverses. In other words, if U_n is one of the elementary gates, then U_n^{-1} should also be one of the elementary gates.

[12] We will not get exactly p^t different unitaries, since more than one sequences might generate the same unitary operation.

[13] This important theorem is sometimes known as the Brylinski theorem after its discoverers, Brylinski and Brylinski [2002]. The authors are married. A simpler version of the proof is given by Bremner et al. [2002].

11.2 Topological Quantum Computing

11.2.1 Hilbert space

With a topological quantum computer, a qubit (or qutrit, etc.) can be formed from multiple anyons which can be put into multiple fusion channels (See chapter 8). For example, with Fibonacci anyons a qubit might be formed from three Fibonacci anyons fusing to τ as shown in Fig. 9.4. With the Ising theory, one might use a cluster of three Ising anyons fusing to σ as as qubit as shown in Fig. 9.5. There are, of course, many more options of how one encodes a qubit in any given theory. For example, in the Ising theory it may be more convenient to work with clusters of four Ising anyons fusing to I as shown in Fig. 11.1.

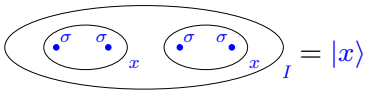


Fig. 11.1 A qubit made from four Ising anyons in an overall fusion channel of I. The two states of the qubit are $x = \psi$ and $x = I$. Note that due to the fusion rules of the Ising theory, if the overall state of the four qubits is I, then if the left two anyons are in state x, the right two must also be in state x. Using a qubit made of 4-anyons has advantages for other topological theories such as $SU(2)_k$ with $k > 4$. See, for example, Hormozi et al. [2009].

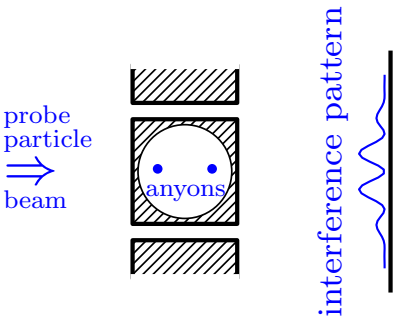


Fig. 11.2 Using Aharanov-Bohm-like interference to measure the fusion channel of two anyons (inside the circle) that are far apart.

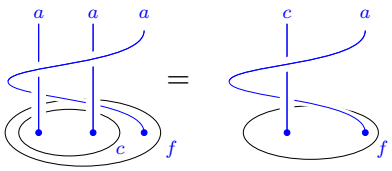


Fig. 11.3 The interference experiment in Fig. 11.2 is equivalent to measuring the phase of wrapping the probe particle (right) around the two test particles. The general expression for the resulting phase would be $\hat{\sigma}_2\hat{\sigma}_1^2\hat{\sigma}_2$, which is dependent on the fusion channel c. (In chapter 13. we will refer to this phase as $R_f^{ca}R_f^{ac}$.)

11.2.2 Measurement (in brief) and initialization

A topological qubit could be measured in several ways, depending on the particular physical system in question. The genereal principle of locality which we introduced in section 8.2 (See in particular Fig. 8.8) gives a good idea how such measurements can (or can't) be done.

Let us suppose, for example, we have two anyons of typs a and we would like to measure their fusion channel. Given the principle of locality, to measure this fusion channel we must perform an operation which is local to both particles, i.e., a measurement that surrounds both.

One way to measure the fusion channel of two anyons is to bring them together to the same point, or at least bring them physically close on a microscopic scale. When two anyons are microscopically close to each other, in essence their wavefunctions mix with each other and in this case measurement of *almost any* nontrivial operator near that location will suffice to distinguish between the different possible fusion channels. For example, one could measure the energy of the two anyons, or the force between them, which would generally distinguish the fusion channels. Note, however, when the anyons are moved macroscopically far apart *no* local operators should be able to distinguish the fusion channels. (We will discuss precisly why this is the case in chapter *** below).

Another way to measure the fusion channel of two anyons would be to leave the two anyons far apart from each other but implement a measurement that surrounds them both — such as Aharanov-Bohm-type interference as shown in Fig. 11.2. Here a test particle wave is split into two partial waves which travel on opposite sides of the anyons to be measured and then reinterfere with each other. This is entirely analogous to the regular Aharanov-Bohm effect (See section 4.1 and Fig. 4.2), where the partial waves travel on opposite sides of a flux and then reinterfere. In the usual Aharonov-Bohm effect, the net phase we measure is the phase of wrapping a single test particle all the way around the central region (See Eq. 4.3). Analogously here we measure the phase of wrapping the probe anyon all the way around the anyons in the central region to measure their fusion channel as shown in Fig. 11.3. As we will discuss in chapter ***, experiments of this sort have been attempted in quantum Hall systems.

Once we know how to measure the state of the anyons in our Hilbert space (and assuming we know how to manipulate our qubits) it is then fairly trivial to initialize the Hilbert space. We simply measure the state of a qubit: If it is in the state we want, we are done. If it is in some other state, we apply the appropriate unitary operation to put it into the desired initial state. We will discuss unitary operations next.

11.2.3 Universal Braiding

The most interesting part of a topological quantum compuatation is the idea that we can apply a controlled unitary operation on our Hilbert space by braiding anyons around each other. The elementary gates of

the system (or elementary unitary operations) are the (counterclockwise) exchanges of two identical anyons, which, in braid group notation, we call $\hat{\sigma}_n$, as well as the inverse (clockwise) exchanges $\hat{\sigma}_n^{-1}$, where $n \in 1, 2, \ldots, (N-1)$ for a system of N identical anyons. Each of these braid operators corresponds to a unitary matrix operating on the Hilbert space.

It turns out that for many types of nonabelian anyon theories, the gate set made up of elementary braiding exchanges is universal in the sense defined in section 11.1.1[14]. For example, braiding is universal for Fibonacci anyons. Similarly $SU(2)_k$ Chern-Simons theory is universal for $k = 3$ and $k > 4$. In fact, among nonabelian anyon theories, theories where braiding is *not* universal are somewhat of an exception. Ising anyons and the closely related $SU(2)_2$ Chern-Simons anyons are two of these nonuniversal exceptions[15]

It turns out that any system of N identical anyons that is capable of universal quantum computation by braiding, is also capabable of universal quantum compuation by *weaving*[16]. Here, what we mean by "weave" is that we fix the positions of $N-1$ of the anyons and only move the one remaining anyon around all the other stationary anyons. An example of a weave is shown in Fig. 11.4. The weaves are a very restricted subset of the possible braids, but still the weaves form a universal set of gates for these anyon systems. This result will be important below in section 11.4.1. (See also exercise 11.5).

In fact, if one is able to measure fusion channels easily[17], it is also possible to implement universal quantum computation just by making many measurements of fusion channels, without physically braiding any particles around any others[18].

[14] This result was shown by Freedman et al. [2002a, b]. These papers are not particularly easy to read for physicists.

[16] This is proven by Simon et al. [2006]. Publication of this work reduced my Erdös number to its current value of 3.

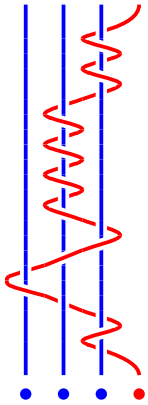


Fig. 11.4 A *weave* is a braid where only one particle moves and all the other particles remain stationary. All the particles in this figure are supposed to be of the same type. The single particle that moves is colored red just for clarity.

[17] Note that measurement schemes of the Aharanov-Bohm type, as in Fig. 11.2, involve braiding a test particle around other particles.

[18] See Ref. Bonderson et al. [2008a].

11.3 Fibonacci Example

As an example, we will focus on the case of Fibonacci anyons, which is potentially the simplest anyon system which is universal for quantum computation.

11.3.1 A Single Fibonacci Qubit

Let us consider a single qubit made of three Fibonacci anyons. We have discussed this several times before in sections 8.2.1, and 9.1 and 10.2.1. To remind the reader, there are three possible states of three Fibonacci anyons which we label $|N\rangle, |0\rangle, |1\rangle$ (See Fig. 8.10) — which represents a qubit (the states $|0\rangle$ and $|1\rangle$) and one additional "noncomputational" state $|N\rangle$ which we will not use for storing quantum information.

[15] $SU(2)_4$ is an interesting case where braiding alone is not universal. However, if we are allowed to go outside of the quantum circuit model and combine braiding with many projective measurements (i.e., not just making one measurement at the end of the computation), then $SU(2)_4$ anyons can impliment universal quantum computation (See Refs. Levaillant et al. [2015]; Cui and Wang [2015]). In fact, the first proposal of a topological quantum computer, by Kitaev in 1997 (published as Kitaev [2003]), described a computation scheme which involved both braiding and projective measurement. A simple discussion of this scheme is given by Preskill [2004] with early extensions of the scheme given by Mochon [2003, 2004].

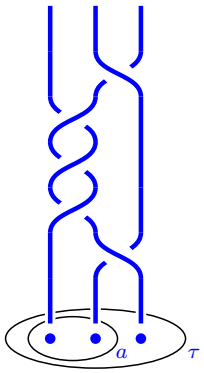


Fig. 11.5 This is the braid written in Eq. 11.6 which gives an approximation of the X-gate on a single qubit made from Fibonacci anyons. As usual, time runs bottom to top. The distance to the taget is **dist** = 0.17

[19]Note that a Z-gate can be implemented exactly as $\hat{\sigma}_1^5$. It is unusual and non-generic that a target can be constructed exactly.

[20]Recall that in comparing Eq. 11.6 to Eq. 11.5 we are not concerned with the overall phase, so we ignore the prefactor of $e^{-3\pi i/5}$ in Eq. 11.6.

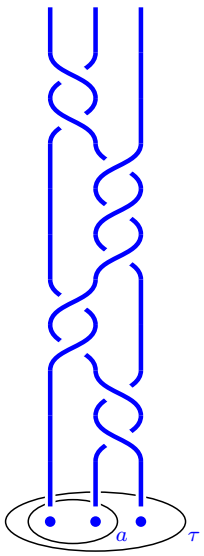


Fig. 11.6 A longer braid gives a more accurate approximation to the desired target X-gate for Fibonacci anyons. This braid has distance to the target, **dist** ≈ 0.08.

We now think about braiding our three anyons. In the braid group on three strand, B_3 (See section 3.3.1), there are two generators, σ_1, exchanging the first two strands counterclockwise, and σ_2, exchanging the second two strands counterclockwise. Any braid of three particles can be constructed as some product of $\sigma_1, \sigma_2, \sigma_1^{-1}$, and σ_2^{-1} in some order as shown, for example, in Fig. 11.5.

The action of these braid operations on the three-dimensional Hilbert space is shown in Eqs. 10.6 and 10.7 which we calculated in section 10.2.1. By multiplying these matrices together, we can figure out how any complicated braid acts on our Hilbert space. In fact for now we are only interested in how the matrices act on the space of the qubit states $|0\rangle$ and $|1\rangle$ and we will return to worry about the $|N\rangle$ state below in section **.

Example of X Gate

We are now interested in the following simple quantum computation problem: Given a particular target unitary operation which we might want to perform on our qubit, how should we move the anyons? I.e., what braid should we do to implement the target operation?

For example, suppose we want to design a braid that impliments an X-gate[19] (just a Pauli σ_x)

$$U_{\text{target}} = X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{11.5}$$

With a very short braid (Shown in Fig. 11.5), we can make a fairly poor approximation to this gate (this braid is the best we can do with only five braid operations) given by

$$U_{\text{approx}} = \hat{\sigma}_2^{-1}\,\hat{\sigma}_1^3\,\hat{\sigma}_2^{-1} \approx e^{-3\pi i/5} \begin{pmatrix} 0.073 - 0.225i & 0.972 \\ 0.972 & -0.073 - 0.225i \end{pmatrix} \tag{11.6}$$

For the approximation given in Eq. 11.6 the phase invariant distance from the target is[20]

$$\mathbf{dist}(U_{\text{target}}\,;\, U_{\text{approx}}) \approx 0.17$$

which is not a great approximation. However, with a longer braid having nine braid operations, shown in Fig. 11.6, one can make a better approximation with a trace distance **dist** ≈ 0.08. If we consider braids that are longer and longer, we can get successively better approximations to the desired target as would be expected from the Kitaev-Solovay theorem discussed in section 11.1.1.

As mentioned in section 11.2.3 it is possible to find braids that are *weaves*, meaning that only a single anyon moves. For completeness, we show a weave in Fig. 11.7 that impliments an X-gate to precision **dist** $\approx .18$. Note that due to the restricted weave form of this braid, a slightly larger number of elementary exchanges are required to reach

roughly the same precision as in Fig. 11.5. As with braids, at least in principle, by using longer weaves one can get as close to the target as we like.

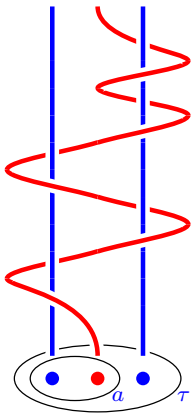


Fig. 11.7 A weave that approximately impliments an X-gate for Fibonacci anyons. Here all three anyons are meant to be identical. The anyon colored red is mobile whereas the other two are kept stationary. The distance to the target is **dist** $\approx .18$. Because we have restricted the form of this braid to be a weave, the braid is longer (has more elementary exchanges) than the one in Fig. 11.5 for roughly the same accuracy.

11.3.2 Topological Quantum Compiling: Single Qubit

Even if there exists a braid that performs a unitary operation that approximates some target operation within some small error distance ϵ, it is a nontrivial task to figure out what that braid is. In other words, how do you know what braid you should implement on your computer in order to perform the desired operation?

The general task of determining which elementary gates should be performed, and in what order, to implement some desired target unitary is known as *quantum compiling*[21]. For a topological quantum computer, the task of designing a braid is therefore known as *topological quantum compiling*. Here we will discuss several approaches to topological quantum compiling in order of their complexity, and their effectiveness.[22]. We continue to focus only on compiling braids for a single Fibonacci qubit. Multi-qubit braids will be discussed in section 11.4 below.

Brute Force Search

If we are willing to accept a fairly poor approximation of our target unitary (a fairly large **dist** between our approximation and the target) we will be able to use a fairly short sequence of our elementary gates (i.e., a short braid). In this case we can consider some maximum gate sequence length t (maximum run time) and search all possible gate sequences of length less than t, choosing the one that best approximates our target. We should expect to achieve a distance to the target that drops exponentially with t, as discussed near Eq. 11.4.

If we are considering a single qubit made of three Fibonacci anyons, our elementary gates are the braid generators $\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_1^{-1}, \hat{\sigma}_2^{-1}$. This means that if we want to search through all braids of length t we have to search roughly 4^t braids. While there are some tricks that allow us to reduce this number somewhat[23], the computational effort[24] will always grow exponentially with the length t. If one wants to make highly accurate approximation of a target unitary with a very small error distance, one can easily obtain a run time t large enough that brute force searching becomes unfeasible.

Kitaev-Solovay Algorithm

Kitaev and Solovay[25] provide us an explicit algorithm to construct very accurate approximations of any desired unitary given a universal set of elementary gates with reasonable (not exponentially growning!) computational effort[24]. The essence of this algorithm is as follows. Let us suppose that by brute-force search we can approximate any unitary

[21] Quantum compiling is analogous to a compiling for a conventional computer, which is the task of starting with a high level programming language and determining which machine-level instructions to impliment at the computer chip level. See Harrow [2001] for a discussion of quantum compiling in general.

[22] The field of topological quantum compiling was started by Bonesteel et al. [2005]. A very nice recent review of the topic as well as discussion of a number of other approaches towards topological quantum compiling is given by Field and Simula [2018].

[23] For example, we might not want to search any braids where σ_i and σ_i^{-1} occur in a row since then they would cancel.

[24] Here we mean the computational effort for the classical computer that we use to design our quantum algortithm!

[25] See again Dawson and Nielsen [2006]; Harrow [2001] for nice discussions of Kitaev-Solovay.

operation to within a distance **dist** $\sim \epsilon_0$ with sequency of elementary gates (elementary braids in the topological case) of length t_0. Let us say that the classical computational time to achieve this is T_0. Now given a target unitary $U^{(0)}_{\rm target}$ that would like to approximate, we start with this brute-force search, and construct our approximation $U^{(0)}_{\rm approx}$ which is accurate to within **dist** $\sim \epsilon_0$. This is our 0^{th} level of approximation of the target. We would next like to repair this approximation with another series of gates to make it more accurate. We thus define

$$U^{(1)}_{\rm target} \equiv [U^{(0)}_{\rm approx}]^{-1}\, U^{(0)}_{\rm target} \quad .$$

If we could find a series of gates that would exactly give us $U^{(1)}_{\rm target}$ we could exactly construct the original objective $U^{(0)}_{\rm target}$ as

$$U^{(0)}_{\rm target} = U^{(0)}_{\rm approx}\, U^{(1)}_{\rm target} \quad .$$

However, it is not obvious that we have any better way to approximate $U^{(1)}_{\rm target}$ than we had to approximate $U^{(0)}_{\rm target}$, so why does this help? The key here is that $U^{(1)}_{\rm target}$ is necessarily close (**dist** $\sim \epsilon_{\mathbf{0}}$) to the identity. We then decompose

$$U^{(1)}_{\rm target} = VWV^{-1}W^{-1}$$

with W and V being unitary operations close to the identity (**dist** $\sim \sqrt{\epsilon_0}$). We then have an amazing result, that if we are able to approxmate V and W to an accuracy ϵ_0 (which we can do here by brute-force search) we will get $U^{(1)}_{\rm target}$ accurate to **dist** $\sim \epsilon_0^{3/2}$. Thus we obtain

$$U^{(0)}_{\rm target} = U^{(0)}_{\rm approx} VWV^{-1}W^{-1} \tag{11.7}$$

accurate to order $\epsilon_0^{3/2}$. The total sequence of gates is now of length $5t_0$ since each of othe factors on the right hand side of Eq. 11.7 is of length t_0. The classical computational effort to achieve this is roughly $3T_0$ since we must search for $U^{(0)}_{\rm approx}$ and V and W.

This scheme can then be iterated to make our approximation even better. The only change it that the next level of approximation, instead of using brute force search to make approximations good to **dist** $\sim \epsilon_0$ we use the entire above described algorithm to make all of our approximations good to **dist** $\sim \epsilon_0^{3/2}$. When $U^{(0)}_{\rm approx}$ and V and W are calculated to order $\epsilon_0^{3/2}$ our new approximation for $U^{(0)}_{\rm target}$ will be an accurate to **dist** $\sim (\epsilon_0^{3/2})^{3/2}$.

The entire scheme can be iterated recursively to any level of accuracy. At the n^{th} level of this approximation, we have a gate sequence of length $5^n t_0$ and an accuracy **dist** $\sim \epsilon_0^{(3/2)^n}$ and the computational effort[24] to achieve this scales as $3^n T_0$.

Thus if we want to achieve some overall accuracy ϵ of our operation,

the gate sequence will be of length

$$t \sim \mathcal{O}\left([\log(1/\epsilon)]^{(\ln(5)/\ln(3/2))}\right) = \mathcal{O}\left([\log(1/\epsilon)]^{3.969\ldots}\right) \tag{11.8}$$

and this requires classical computation time

$$T \sim \mathcal{O}\left([\log(1/\epsilon)]^{(\ln(3)/\ln(3/2))}\right) = \mathcal{O}\left([\log(1/\epsilon)]^{2.710\ldots}\right). \tag{11.9}$$

While this algorithm produces gate sequences that are longer than one obtains with brute force searching (which produces gate sequence lengths as in Eq. 11.4) for the same desired accuracy ϵ, it has the advantage that it is computationally feasible[24] for much smaller values of ϵ and can therefore produce more accurate results.

Galois Theory Optimal Compiling:

A rather remarkable scheme for quantum compiling was developed in Kliuchnikov et al. [2014] based on ideas from Galois theory[26] . While we cannot review Galois theory here, nor can we even do justice to the details of the algorithm, we can nonetheless discuss some of structure of the problem that makes this approach possible.

It turns out that any unitary that can be constructed by braiding three Fibonacci anyons can be written (up to a phase) in the form

$$U(u, v, k) = \begin{pmatrix} u & v^* \omega^k \phi^{-1/2} \\ v\phi^{-1/2} & -u^* \omega^k \end{pmatrix} \tag{11.10}$$

where $\phi = (1 + \sqrt{5})/2$ is the golden mean, k is an integer, $\omega = e^{2\pi i/10}$, and

$$|u^2| + \phi|v|^2 = 1 \tag{11.11}$$

where u and v come from the so-called ring of cyclotomic integers $\mathbb{Z}[\omega]$, which means that

$$u = \sum_{i=0}^{3} a_i \omega^i \qquad v = \sum_{i=0}^{3} b_i \omega^i \tag{11.12}$$

with coefficients a_i and b_i all being integers. The fact that the unitaries that can be generated by braiding take a very restricted mathematical form is, in fact, a generic property of all anyon theories[27], although the particular form taken depends on the particular anyon theory.

Further, given values of u, v, and k a relatively simple algorithm is provided that finds a braid[28] that results exactly in this unitary, where length of the braid is no longer than

$$t \sim \log\left(|\sum_{i=0}^{3} a_i \omega^i|^2 + |\sum_{i=0}^{3} a_i \omega^{3i}|^2\right)$$

This procedure is known as *exact synthesis* as it constructs exactly the

[26] Évariste Galois was undoubtedly one of the most interesting and brilliant mathematicians of all time. Being politically active in an era shortly after the French revolution, he spent a decent fraction of his short adult life in prison. His mathematical works (some written while in prison) opened up vast new fields of research. He died at age 20 in a duel.

[27] This is due to the fact that the F and R matrices of an anyon theory live in a particular so-called Galois extension of the rationals — meaning that only certain irrational factors can show up in any mathematical expression. This fact can be used to prove various statements about what type of operations can or cannot be done exactly by braiding. See for example Freedman and Wang [2007].

[28] Kliuchnikov et al. [2014] also provide a similar algorithm for generating *weaves*. See section 11.2.3.

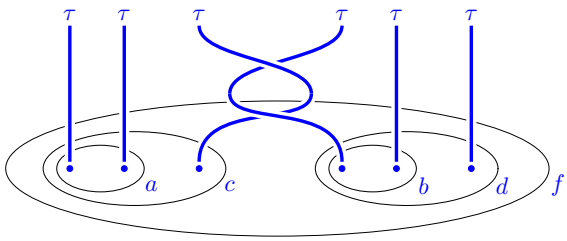


Fig. 11.8 The braid shown here between two Fibonacci qubits entangles the two qubits but also results in leakage error. When we use three Fibonacci anyons as a qubit, we set the overall fusion channel of the three to be τ, so $c = d = \tau$ in this figure. The quantum information is stored in the quantum numbers a and b. The shown braid results in some amplitude ending up in the noncomputational space where eitehr c or d is I rather than τ.

desired $U(u, v, k)$ as a series of elementary braid operations.

The remainder of the algorithm is to find values of u, v, k (with u and v of the form in Eq. 11.12 with integer coefficients) so that Eq. 11.10 approximates any given target unitary. This task can exploit established methods from algebraic number theory. The interested reader is referred to Kliuchnikov et al. [2014].

The end result of this approach is an algorithm that, although it does not find the absolute optimal braid[29], it nonetheless is *asymptotically optimal* in the sense that it produces braids of length

$$t \sim \mathcal{O}(\log(1/\epsilon))$$

as in Eq. 11.4. Further, the computational time[24] to achieve this scales only as $T \sim \mathcal{O}(\log(1/\epsilon)^2)$. Using this type of approach, it is easily possible to generate braids with error distances of order 10^{-100} or even better, and these braids are longer than the absolute optimal braid by only a factor of order unity.

[29]The "optimal braid" is the one that would be found by brute force search if one had the exponentially enormous computational power necessary to find it.

11.4 Two-Qubit Gates

Having studied single qubit operations, we now turn to a brief discussion of two-qubit gates[30]. As mentioned in section 11.1.1, the Brylinski theorem tells us that to have a universal quantum computer, we need only have single qubit rotations along with any entangling two qubit gate. To construct such an entangling two qubit gate we will need to have a braid that physically entangles the world lines of the anyons comprising the two qubits such as the example shown in Fig. 11.8.

However there is a crucial complication with braiding anyons between qubits. If we perform a braid such as that shown in Fig. 11.8, the fusion channel of the anyons comprising each of the qubits (quantum numbers c and d in the figure) are not preserved (see the discussion of locality in section 8.6) and this means that amplitude can *leak* into the noncomputational space.

To be more explicit for the Fibonacci case, recall that we encode our qubits ($|0\rangle$ or $|1\rangle$) by using three Fibonacci anyons in overall fusion

[30]Here we are constructing a two qubit unitary operation, which we will call a two-qubit gate, from our elementary gates —- the elementary braid operations.

channel τ (See Fig. 8.10). The fusion channel of the three anyons to I is termed non-computatational $|N\rangle$ and is not used for computation. If some of the amplitude of the wavefunction ends up in this noncomputational space, it is called *leakage error*, and only very small quantities of such leakage errors can be tolerated for any realistic computation. Braids like the one shown in Fig. 11.8 always produce some amplitude of noncomputational states. The problem of leakage error in two-qubit gates is not special to Fibonacci anyons, but is in fact a generic property of all anyon theories that have universal braiding[31].

While we cannot completely eliminate leakage, we can in principle design entangling gates with arbitrarily small (albeit non-zero) leakage. Such braids with low leakage error do exist, but finding them is highly nontrivial. Inconveniently, the Hilbert space of six Fibonacci anyons, as in Fig. 11.8 is 13 dimensional[32]. Searching such a large space for particular unitaries with low leakage is numerically unfeasible. We thus need a more clever way to design braids with low leakage.

In designing any computation, it is almost always advantageous to simplify the desired task into smaller tasks that can be addressed one at a time. This "divide and conquor" approach will allow us to tackle the job of desiging two-qubit gates. In the next section we will give an example of how entangling gates with negligible leakage can be designed.

[31] Ainsworth and Slingerland [2011] show that it is not possible to design completely leakage free gates for any universal anyon theory, and leakage can only be make approximately zero.

[32] This space is subdivided into an 8 dimensional subspace with $f = \tau$ and a 5 dimensinal subspace with $f = I$. No braiding of these six anyons will change the f quantum number. Note, however, that gates must have low leakage independent of the value of f.

11.4.1 Controlled Gates

In quantum computation it is often very convenient to use entangling gates which are so-called controlled U-gates, or $C(U)$ where U is a 2-dimensional unitary matrix. A controlled U-gate acts on two qubits such that one qubit (the "target" qubit) is acted on with a 2-dimensional unitary operator U if and only if the other qubit (the "control" qubit) is in the $|1\rangle$ state, whereas the control qubit remains unchanged:

$$C(U): \quad \begin{cases} |0\rangle \otimes |0\rangle & \rightarrow & |0\rangle \otimes |0\rangle \\ |1\rangle \otimes |0\rangle & \rightarrow & |1\rangle \otimes |0\rangle \\ |0\rangle \otimes |1\rangle & \rightarrow & (U|0\rangle) \otimes |1\rangle \\ |1\rangle \otimes |1\rangle & \rightarrow & (U|1\rangle) \otimes |1\rangle \end{cases} \tag{11.13}$$

Thus the first qubit here is being controlled by the second qubit[33]. A very commonly used example of a controlled gate is the case of $U = X$ (See Eq. 11.5) which we call a controlled-X, or more often a controlled-NOT (or CNOT) gate.

The key to our construction of controlled gates[34] is the locality principle of section 10.1.1. If we are given a cluster of 2 anyons which are the τ fusion channel (for example, set $c = \tau$ in Fig. 10.7) and we braid it around some other anyons, this will have the same effect as if we just braided a single τ around the other anyons. However, if the cluster of 2 anyons is in the trivial (or I) fusion channel, then braiding this cluster never does anything, as braiding the vacuum particle always is trivial. Thus we can see that the effect of the braid is "controlled" by the fusion channel of the two anyons.

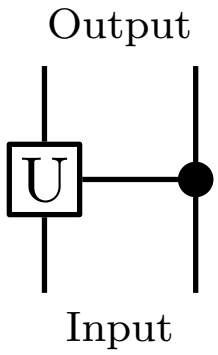


Fig. 11.9 Typical notation for a controlled unitary gate $C(U)$. The second qubit controls the first.

[33] As always in quantum mechanics the operation acts linearly on superpositions

[34] The constructions discussed here were introduced in Bonesteel et al. [2005].

Controlled $\hat{\sigma}_2^2$ gate

Consider the construction shown in Fig. 11.10. On the far left of this figure, we have shown a *weave*, meaning only a single anyon, the one drawn in red, moves and the other two anyons remain stationary (see the discussion by Fig. 11.4). This weave has been designed to have approximately the same effect as if the two blue anyons are wrapped around each other (exchanged twice counterclockwise), i.e, $\hat{\sigma}_1^2$ as shown in the figure. For the particular weave shown, the distance to the target $\hat{\sigma}_1^2$ is **dist** $\approx .12$. We could make the approximation of $\hat{\sigma}_1^2$ more accurate by using a longer weave using any of the compiling methods discussed in section 11.3.2 above. Note that the equivalance between the weave on the far left and σ_1^2 is true as a 3×3 matrix acting on the full three dimensional Hilbert space spanned by three Fibonacci anyons (i.e., on the space spanned by $|0\rangle, |1\rangle, |N\rangle$, not just $|0\rangle, |1\rangle$).

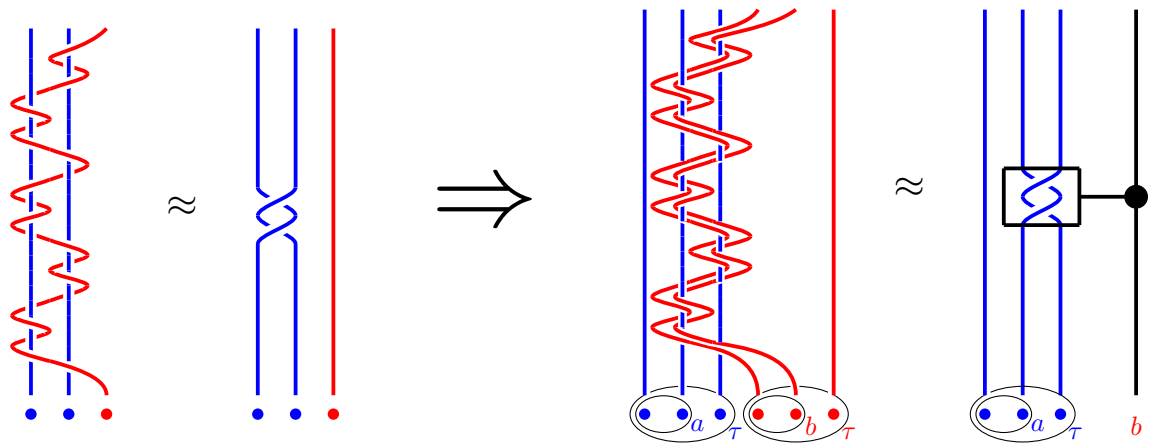


Fig. 11.10 Construction of a controlled gate using Fibonacci anyons. The weave on the far left is designed to have almost the same effect on the Hilbert space as the braiding (two counterclockwise exchanges) of the two blue particles as shown. Using a longer weave one can more closely approximate the braiding of the two blue particels. On the right, we have a system of 6 anyons representing two qubits. The right (red) we take the cluster of two red anyons, labeled b

Now consider the braid of six anyons on the right of Fig. 11.10 representing two qubits — the right (red) anyons are the control qubit and the left (blue) anyons are the target qubits. We group the two red anyons in fusion channel b and we move them around as a single unit to form the same weave as shown on the far left (here using the two red anyons and the right two anyons of the blue qubit). If these two red anyons are in the vacuum fusion channel $b = I$, then this braiding has no effect on the Hilbert space (braiding of the vacuum particle is always trivial). On the other hand, if the two anyons are in fusion channel $b = \tau$ then this braid is equivalent to moving a single τ particle through exactly the same weave as on the far left, thus having the same effect as exchanging the two right-most blue anyons twice counterclockwise. We have thus

constructed a controlled operation, $C(\hat{\sigma}_2^2)$ which is notated on the far right of the figure in a manner analogous to Fig. 11.9: The operation implemented on the blue qubit is (approximately) a full braiding of the right two blue strands, if and only if the right qubit (b) is in the τ or $|1\rangle$ or τ state.

A crucial feature of this construction is that, to the extent that the weave we use accurately approximates $\hat{\sigma}_1^2$, the resulting construction leads to no leakage error. The right hand qubit (b) is completely unchanged (hence not creating leakage of this qubit), and the effect on the left hand qubit (a) is designed to be equivalent to just braiding two of the blue anyons — which does not create leakage either.

Controlled U-gate

With a bit more work, we can in fact make any controlled $C(U)$ gate for an arbitrary two dimensional unitary U, as in Fig. 11.9 (up to an overall phase as discussed in section11.1.1).

First let us discuss the so-called *injection weave* described in Fig. 11.11. An ideal injection weave is meant to leave the Hilbert space unchanged (it only applies an identity matrix). However, it has the nontrivial effect of rearranging the three strands comprising a qubit. As shown in the diagram, the injection weave moves the red strand from the far right at the bottom to the far left at the top. As discussed in section 11.3.2 we can more precisely approximate the ideal injection by using a longer weave.

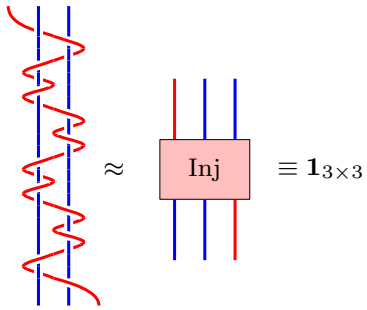


Fig. 11.11 An approximate injection weave is shown on the left. A perfect injection has no effect on the Hilbert space (it applies the identity matrix to the Hilbert space) but moves the red strand from the right to the left of the three anyons. The distance to the target for this particular weave is **dist** = 0.09. With a longer weave one can more accurately approximate a perfect injection.

We now construct the braid shown on the left of Fig. 11.12. As in Fig. 11.10 we group together the two red anyons in fusion channel b and we move them around as a group. These two anyons are first put through an injection weave with the right most two blue anyons. This moves the group of two anyons into the middle position of the right qubit. A weave to implement an arbitrary unitary U is then implemented on the three strands furthest left, treating the two red strands grouped together as a single strand. Finally, thie injection weave is inverted to bring the two red particles back to their original position. The braid constructed in this way will implement a controlled U gate $C(U)$, as shown using the notation of Fig. 11.9 on the right of Fig. 11.12: The left (blue) qubit (a) has the unitary U applied to it, if and only if the right qubit (b) is in the $|1\rangle$ or τ state. If we choose U to be an X gate, such that the necessary weave in the middle step is a weave like that shown in Fig. 11.7, we obtain a $C(X)$ or controlled NOT gate (CNOT).

To understand this procedure we realize that the only two anyons that are moved in this procedure are the two red anyons in state b, and these two are moved as a group. As in Fig. 11.10 if these two anyons are in fusion channel $b = I$ (or $|0\rangle$) then the Hilbert space is left unchanged. However, if $b = \tau$ (or $|1\rangle$) then there will be an effect on the blue qubit — hence we have a controlled rotation. Let us now consider this case when $b = \tau$, so that we should think of the two red strands as being a single τ strand. The injection weaves are designed to have no effect on

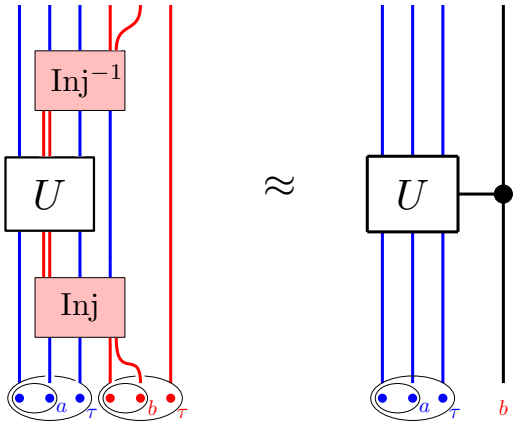


Fig. 11.12 Construction of a controlled U gate $C(U)$ with Fibonacci anyons. The two red anyons in state b are moved as a group and all other anyons are kept stationary. If $b = I$ or $|0\rangle$, then the weave has no effect on the Hilbert space. If $b = \tau$ or $|1\rangle$ then this weave implements a U rotation on the left (a) qubit.

the Hilbert space — their only effect is to move the red double strand inside of the blue qubit. The unitary rotation U is thus only nontrivial effect on the Hilbert space.

Chapter Summary

- Summary Item 1
- Summary Item 2 etc

Further Reading

MOst of the same ideas can be generalzied for other anyon systems.

Freedman et al. [2002a] Freedman et al. [2002b] for the initial work showing SU(2)k is universal

Mochon [2003] Mochon [2004] for universal quantum computing with topological superconductor. (This was slightly after freedman)

Bonderson et al. [2008b] for interferometry

Bonderson [2009] for splitting of topological degeneracy

Bonderson et al. [2008a] for measurement only tqc.

Field and Simula [2018] Simula review

Bonesteel et al. [2005] Original Compiling Fib anyons

Hormozi et al. [2007] Quant compiling PRB

Hormozi et al. [2009] Compiling RR states

Simon et al. [2006] One mobile particle

Nielsen and Chuang [2000] Nielson and Chuang

Harrow [2001] Harrow's thesis.

Brylinski and Brylinski [2002] Brylinksi theorem

Kliuchnikov et al. [2014] Galois theory

Maybe cite original Solovay and Kitaev article.

Levaillant et al. [2015] Cui and Wang [2015] For SU(2)4

.

Exercises

Exercise 11.1 Ising Nonuniversality

The braiding matrices for Ising anyons are given by Eqs. 10.10 and 10.11. Demonstrate that any multiplciation of these matrices, and there inverses will only produce a finite number of possible results. Thus conclude that Ising anyons are not universal for quantum compuation. Hint: write the braiding matrices as $e^{i\alpha}U_i$ where U_i is unitary with unit determinant, i.e., is an element of $SU(2)$. Then note that any $SU(2)$ matrix can be thought of as a rotation $\exp(i\,\hat{n}\cdot\boldsymbol{\sigma}\,\theta/2)$ where here θ is an angle of rotation $\hat{n}$ is the axis of rotation and $\boldsymbol{\sigma}$ is the vector of Pauli spin matrices.

Exercise 11.2 Brute Force Search

Given the braid matrices for Fibonacci anyons in Eq. 10.6 and 10.7, write a computer program for brute-force searching braids up to length 10.

Ignoring the noncomputational state $|N\rangle$, and ignoring the overall phase as usual, determine the closest approximation to the Hadamard gate

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

Partial Answer: A braid of length 10 exists with phase-invariant distance to target **dist** ≈ 0.084

Exercise 11.3 Scaling of Kitaev-Solovay Algorithm

Given the discussion just above Eq. 11.8, prove Eqs. 11.8 and 11.9.

Exercise 11.4 About the Injection Weave

One might wonder why we choose to work with an injection weave in Fig. 11.11 which moves the red strand from the far right at the bottom all the way to the far left on the top. Show that for three Fibonacci anyons, there does not exist any injection weave that moves the (red) strand from the far right on the bottom to the *middle* on the top, even up to an overall phase. I.e., show that no weave exists starting on the bottom far left ending in the middle on the top whose effect on the three dimensional Hilbert space is $e^{i\phi}\mathbf{1}_{3\times 3}$ for any phase ϕ.

Exercise 11.5 Universal Weaving and the Injection Weave

Consider injection weaves as described in Fig. 11.11. Let us assume that we can construct an injection weave of arbitrary precision. Given such an (approximately) perfect injection weave show that for any number of anyons $N > 3$, a weave can be constructed that performs the same unitary operation on the Hilbert space as any given braid. A more general mathematical proof of the universality of weaving is also given in Simon et al. [2006].

Part III

Anyon Diagramatics (in detail)

12

Planar Diagrams[1]

One of our objectives is to come up with some diagrammatic rules (somewhat analogous to those of the Kauffman bracket invariant) which will allow us to evaluate any diagram of world-lines (i.e, a labeled link, possibly now including diagrams where particles come together and fuse, or split apart) and get an output which is a complex number as desired in Fig. 7.1. In chapters 8-10 we have been putting together some of the necessary pieces for these diagrammatic rules. Here we will begin to formalize our diagrammatic algebra a bit more precisely[2]. While we try to physically motivate all of our steps, in essence the rules of this chapter can be taken to be axioms of the diagrammatic algebra.

In this chapter we will focus only on planar diagrams — i.e., we do not allow lines to cross over and under each other forming braids. We can roughtly think of such planar diagrams as being 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 they split apart. An example of a planar fusion diagram is shown in Fig. 12.1. 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 that we will discuss. These two normalization conventions are useful in different contexts. We will start with a more "physics" oriented normalization in this chapter but we switch to a more topologically oriented normalization in chapter 14 and in later chapters.

We start by briefly reviewing some of the notions introduced in chapters 8-9: We assume a a set of particle types $a, b, c, \ldots$ which we will draw as labeled lines with arrows in our diagrammatic algebra. This set of particles includes a unique identity or vacuum particle I, which may be drawn as a dotted line, or or may not be drawn at all since it corresponds to the absence of any particles. Each particle type has a unique antiparticle denoted with an overbar ($\bar{a}$ for the antiparticle of a). As we discussed in section 8.1, if we reverse the arrow on a line we turn a particle into its antiparticle. If a particle is its own antiparticle we do not draw an arrow on its line.

Fusion rules are given by the matrices N_{ab}^c having the properties discussed in section 8.3. We will also assume a consistent[3] set of F-matrices as discussed in chapter 9.

[1] This chapter through chapter 15 develop the diagrammatic algebra in some detail. Those who would like a brief and easier (albeit not as general) introduction to diagrammatic algebra should go straight to chapter 16.

[2] Formally, some of the mathematical structure of planar diagrams was introduced in section 8.6. The rules we are defining in this chapter is known as a *unitary fusion category* to mathematicians.

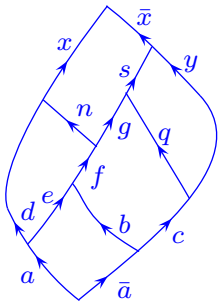


Fig. 12.1 A planar fusion diagram starting and ending at the vacuum.

[3] The word "consistent" here means that the F-matrices satisfy the pentagon Eq. 9.7.

12.1 Diagrams as Operators

If, like Fig. 12.1, a diagram starts at the bottom from the vacuum and ends at the top with the vacuum, we interpret that diagram to represent a complex number, or an amplitude. However, we will also consider diagrams that have "loose ends" (lines sticking off the top or bottom of the page) meaning that they may not begin or end with the vacuum[4]. We can view these diagrams with loose ends as being a sub-diagram of a larger diagram that begins and ends in the vacuum. However, it is also useful to give such diagrams quantum mechanical meaning in their own right.

[4]Many of the diagrams we have drawn (such as Fig. 8.1 or Fig. 9.1) have not started at the bottom with the vacuum or ended at the top with vacuum.

Our convention is that when we draw a diagram with world-lines that end pointing upwards we should view these particles as kets (independent of the direction of any arrow drawn on the world-line). If world-lines end pointing downwards, we mean them to be bras. Many diagrams will have world-lines that point both up and down, in which case 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. The lines coming in from the bottom are thus in the bra part of the operator and the lines pointing out the top are the ket part of the operator[5]. If we consider, for example, diagrams with M_{in} incoming lines from the bottom and M_{out} lines going out the top, we can write a general operator[6] as

$$\text{Operator} = \sum_{n,m,q} C_{n,m,q} \; |n, M_{\text{out}}; q\rangle\langle m, M_{\text{in}}; q| \tag{12.1}$$

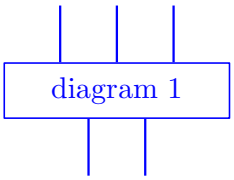
[5]Analogous to some of the ideas of chapter 7, the bras and kets are meant to be contracted together with bras and kets from other diagrams, pasting together such operators to assemble a picture with no loose ends like Fig. 12.1 which starts and ends in the vacuum.

[6]The only constraint on this operator is that it conserves the total quantum number (or "charge"). One could imagine operators that do not conserve total quantum number. Such operators would be nonphysical and are also outside of what we can express with diagrams.

An example of such an operator is shown diagrammatically in Fig. 12.2 with two incoming and three outgoing lines. In Eq. 12.1 the states $|n, M_{\text{out}}; q\rangle$ are an orthonormal complete set of states of M_{out} particles where all the particles together fuse to the quantum number q; and similarly the states $|m, M_{\text{in}}; q\rangle$ are an orthonormal complete set of states of M_{in} particles where all the particles together fuse to the quantum number q. The value of the coefficients $C_{n,m,q}$ depend on the details of the diagram being considered. The fact that the operator is necessarily diagonal in the variable q means that the total quantum number of all of the incoming particles must be the same as the total quantum number of all the outgoing particles (i.e., they fuse to the same overall charge). This conservation of overall quantum number is a reflection[7] of the locality principle of section 8.2.

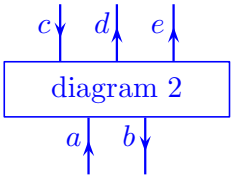
[7]We do not need an axiom for total quantum number conservation, as this will arise as a result of the other rules we introduce in this chapter.

Generally in a diagram, lines will be labeled with particle types and (if the particle is not self-dual) arrows. We have not labeled the incoming and outgoing lines in Fig. 12.2 with the assumption that these labels and arrows occur inside the hidden box. However, it is sometimes useful to reinstate these labels as in Fig. 12.3. As we will discuss in more detail in section 12.2.1 a label restricts the quantum number of the corresponding line



$$= \sum_{n,m,q} C^{(1)}_{n,m,q} \; |n,3;q\rangle\langle m,2;q|$$

Fig. 12.2 Diagram 1, representing an arbitrary diagram (or linear combination, i.e., weighted sum, of diagrams), is understood as part of a larger diagram, and is interpreted as an operator. Incoming lines from the bottom correpond to bras and outgoing lines towards the top correspond to kets. The states $|n,3;q\rangle$ are a complete set of states for 3 particles where all the particles together fuse to the quantum number q. Similarly, the states $|m,2;q\rangle$ are a complete set of states for 2 particles where all the particles together fuse to the quantum number q, The superscript on $C^{(1)}_{n,m,q}$ indicates that these constants correspond to the particular "diagram 1" in the box. The total quantum number q of all the particles is conserved by the operator due to the locality principle from section 8.2.



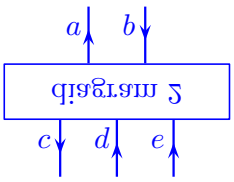
$$= \sum_{n,m,q} C^{(2)}_{n,m,q} \; |n,3;q\rangle\langle m,2;q|$$

Fig. 12.3 In a figure with labeled incoming and outgoing lines, the quantum numbers on these lines are fixed, as compared to Fig. 12.2 where the diagram may have a superposition of quantum numbers on the external lines.

We now introduce an important diagrammatic principle.

> **Hermitian Conjugation:** Reflecting a diagram around a horizontal axis and then reversing the direction of all arrows implements Hermitian conjugation

For example, reflecting Fig. 12.3 and then reversing the arrows on all lines results in the Hermitian conjugate diagram



$$= \sum_{n,m,q} [C^{(2)}_{n,m,q}]^* \; |m,2;q\rangle\langle n,3;q|$$

Fig. 12.4 Flipping the diagram in Fig. 12.3 results in the Hermitian conjugatate. The coefficients $C^{(2)}_{n,m,q}$ in Fig. 12.3 are complex conjugated to obtain $[C^{(2)}_{n,m,q}]^*$ here.

It is crucial that when we turn a bra into a ket (reflecting the diagram and then reversing the arrows), down-pointing arrows remain down-pointing and up-pointing arrows remain up-pointing (Note, for example, that the arrow on a is pointing up both in Fig. 12.3 and 12.4).

Diagrams which start from the vacuum at the bottom are an important special case. When there are no incoming lines at the bottom of a

diagram the expression become

$$\text{ket} = \sum_{n,q} C_{n,\varnothing,q} |n, M_{\text{out}}; q\rangle \tag{12.2}$$

which we can also interpret as an "operator" which accepts the vacuum as an input at the bottom and gives a ket as an output at the top. The symbol $\varnothing$ here means that the m index used in Fig. 12.2 and Eq. 12.1 is just the empty set (nothing summed over), or equivalently that the diagram starts from the vacuum. An example of such a diagram is shown in Fig. 12.5.

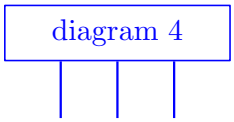
$$= \sum_{n,q} C^{(3)}_{n,\varnothing,q} \; |n, 3; q\rangle$$

Fig. 12.5 A diagram with no incoming lines at the bottom is interpreted as a ket.

Similarly we can consider diagrams which end in the vacuum at the top. When there are no outgoing lines at the top of a diagram we have

$$\text{bra} = \sum_{m,q} C_{\varnothing,m,q} \langle m, M_{\text{in}}; q| \tag{12.3}$$

which is an operator that accepts a ket as an input and gives a complex number as an output. An example of such a diagram is shown in Fig. 12.6.



$$= \sum_{m,q} C^{(4)}_{\varnothing,m,q} \; \langle m, 3; q|$$

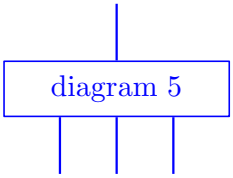
Fig. 12.6 A diagram with no outgoing lines at the top is interpreted as a bra.

If diagram 3 happens to be the reflection of diagram 4 around a horizontal axis with all arrows reversed, then these two diagrams are hermitian conjugates of each other and $C^{(4)}_{n,\varnothing,q} = [C^{(3)}_{\varnothing,n,q}]^*$.

12.1.1 Stacking operators

Stacking operators on top of each other contracts bras with kets in the natural way[8]. For example, if we define the operator, diagram 5, as in

[8]The observant reader will see similarities between this stacking procedure and the stacking of manifolds with boundary discussed in chapter 7. These similarities are not a coincidence!



$$= \sum_{n,m,q} C^{(5)}_{n,m,q} \; |n, 1; q\rangle\langle m, 3; q|$$

Fig. 12.7 Another example operator

Fig. 12.7, we can then stack diagram 5 (Fig. 12.7) on top of diagram 1 (Fig. 12.2) to obtain Fig. 12.8. The resultant operator, diagram 6, on the right is given by

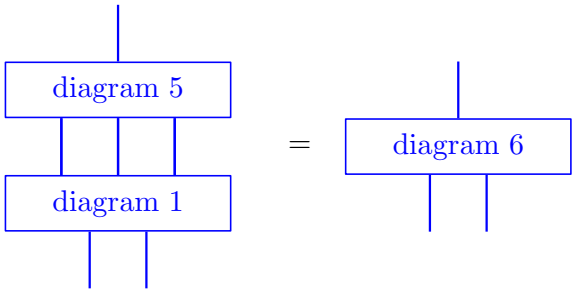


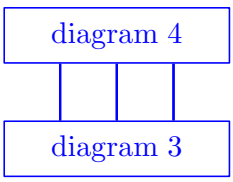
Fig. 12.8 Stacking operators on top of each other to create new operators.

$$\begin{aligned}\text{Operator} &= \left(\sum_{n,m',q} C^{(5)}_{n,m',q}\,|n,1;q\rangle\langle m',3;q|\right)\left(\sum_{n',m,q} C^{(1)}_{n',m,q}\,|n',3;q\rangle\langle m,2;q|\right)\\ &= \sum_{n,m,q}\left(\sum_{n'} C^{(5)}_{n,n',q}C^{(1)}_{n',m,q}\right)|n,1;q\rangle\langle m,2;q|\end{aligned}$$

where we have used the orthonormality of the states $|n',3;q\rangle$ to generate a Kronecker delta $\delta_{m',n'}$. Thus diagram 6 can be written in the usual form of Eq. 12.1 with constants

$$C^{(6)}_{n,m,q} = \sum_{n'} C^{(5)}_{n,n',q}C^{(1)}_{n',m,q}$$

A particularly important case is that of stacking a bra diagram on top of a ket diagram which generates a scalar. For example, stacking the bra diagram 4 on top of the ket diagram 3 generates the usual scalar inner product as shown in Fig. 12.9. This fits with our claim at the beginning of this chapter that a diagram that starts and ends in the vacuum should correspond to a complex amplitude.



$$\begin{aligned}&= \left(\sum_{m,q} C^{(4)}_{\varnothing,m,q}\langle m,3;q|\right)\left(\sum_{n,q} C^{(3)}_{n,\varnothing,q}|n,3;q\rangle\right)\\ &= \sum_{m,q} C^{(4)}_{\varnothing,m,q}C^{(3)}_{m,\varnothing,q}\end{aligned}$$

Fig. 12.9 Stacking a bra operator on top of a ket operator generates a scalar. We have used orthonormality of the kets $|n,3;q\rangle$ on the right hand side.

12.2 Basis of States

In our definition of an operator (Eq. 12.1) we invoked the existence of a complete orthonormal basis of states $|n, M; q\rangle$ for M particles having total quantum number q. We now would like to specify some details of this basis.

12.2.1 One Particle

We begin by considering a single particle at a time. For a single particle, an orthogonal complete basis is given by the different particle types[9] $|a\rangle$ (including the vacuum $|I\rangle$). We denote a projector onto a particular particle type as a simple labeled straight line as shown in Fig. 12.10. The vacuum can be drawn as a dotted line, or may not be drawn at all.

[9] In keeping with the notation of Fig. 12.2 the state $|a\rangle$ should be notated $|a, 1; a\rangle$ to indicate a single line, but here we use just $|a\rangle$ for simplicity.

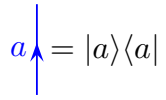


Fig. 12.10 A labeled straight line is just an projector onto the particle type.

Since the different particle types are assumed orthonormal $\langle a|b\rangle = \delta_{ab}$. Applying two such projectors in a row diagrammatically gives the identity shown in Fig. 12.11.

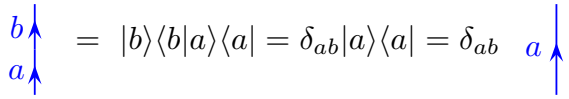


Fig. 12.11 Orthogonality of projection operators.

This identity exemplifies the more general rule shown in Fig. 12.12 which also agrees with the fact that the operators in Eq. 12.1 are diagonal in the overall quantum number q. Again this is simply a reflection of the locality, or no-transmutation, principle[10] of section 8.2 (See in particular Fig. 8.7).

[10] As mentioned in note 7 from earlier in this chapter, this principle is not an axiom of our diagrammatics, but rather can be derived from the other rules we introduce in this section. See exercise 12.2.

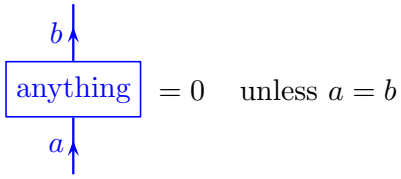


Fig. 12.12 The locality, or no-transmutation, principle as in Fig. 8.7.

Since we assume the set of particle types is complete, the identity operator is given by the sum over all particle types as in Fig. 12.13 where the sum includes the vacuum particle. We represent the identity

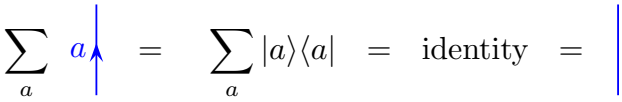
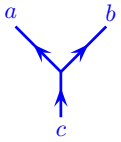


Fig. 12.13 The completeness relation for single lines.

operator on the right in Fig. 12.13 as a straight unlabeled line. This is convenient since it allows us to extend labeled lines by appending unlabeled lines.

12.2.2 Two Particles

Let us now move on to the case of two particles. As discussed in chapter 8, to fully describe the state of two particles, we need to give the quantum number (particle type) of each particle *and* the fusion channel between the two particles. We thus draw the state of two anyons with a vertex diagram[11] as shown in Fig. 12.14.

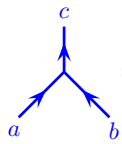


$$= \quad |a, b; c\rangle$$

Fig. 12.14 Particles a and b have fusion channel c.

The notation of the ket[12] $|a, b; c\rangle$ means that the total quantum number of particles a and b is c (or a and b fuse[13] to c). If $N_{ab}^c = 0$, i.e., if the diagram is a disallowed fusion, then the value of the diagram is zero. The set of states $|a, b; c\rangle$ for all possible a, b, c is assumed to form an orthonormal complete set of states for two anyons. Note in particular that for $a \neq b$ the ket $|a, b; c\rangle$ is orthogonal to $|b, a; c\rangle$ — i.e., in our planar diagram algebra, it matters which particle is to the left and which is to the right.

The Hermitian conjugate of the vertex ket Fig. 12.14, the corresponding bra, is shown in Fig. 12.15.



$$= \quad \langle a, b; c|$$

Fig. 12.15 This is the bra which is the Hermitian conjugate of the ket in Fig. 12.14

The fact that the diagram for the bra looks like the ket upside-down is in accordance with our general principle of Hermitian conjugation[14] introduced in section 12.1 (See the discussion near Fig. 12.4).

To take inner products between a bra (like Fig. 12.14) and a ket (like Fig. 12.15) we simply stack the bra on top of the ket, in accordance with section 12.1.1, to produce the diagram[15,16] shown in Fig. 12.16.

The fact that we obtain delta functions on the right is equivalent to the statement that the kets $|a, b; c\rangle$ form an orthonormal set. The normalization of Fig. 12.16 (i.e, that one gets kronecker deltas on the right and no numerical constants) is our *physics normalization*. This normalization will be changed in chapter 14.

Note that the first two delta functions $\delta_{aa'}$ and $\delta_{bb'}$ in Fig. 12.16 can be interpreted as a result of Fig. 12.11 (the lines are angled instead of vertical, but this does not change their meaning). As a result, the

[11] In cases where the fusion multiplicity $N_{ab}^c > 1$ we must also add an index $\mu \in 1 \ldots N_{ab}^c$ at the vertex as in Eq. 8.16, and we would write the ket as $|a, b; c, \mu\rangle$. We suppress this additional index here for simplicity. It is reinstated in section 12.5.

[12] In the notation of Fig. 12.2 the state $|a, b; c\rangle$ might be notated $|(a, b), 2; c\rangle$ to indicate there are two outgoing lines. If we wanted to emphasize that there is one incoming line and two outgoing lines we might write $|a, b; c\rangle\langle c|$ instead. Here we use abbreviated notation.

[13] More properly for Fig. 12.14 we should say that a and b split from c, whereas in Fig. 12.15 we should say that a and b fuse to c. Most of the time people are careless in distinguishing fusing and splitting.

[14] In section 12.1 we treated the statement that flipping the diagram gives Hermitian conjugation as an axiom. However, one could instead treat Fig. 12.15 as the axiom and build up the general principle from only this statement.

[15] Again, if $N_{ab}^c > 1$ there are additional indices μ at the vertices and the kets are orthonormal in these indices as well. See note 11 above, and section 12.5.

[16] This inner product between bra and ket does not give a scalar but rather gives a scalar times a c particle line. This is because the ket, Fig. 12.14 is actually an operator that takes an incoming single line as input and gives two lines as output. (And conversely with the bra Fig. 12.15). See also the comment on notation in the above note 12.

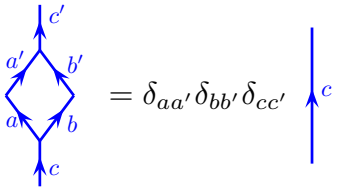


Fig. 12.16 The inner product between the bra in Fig. 12.14 and a ket in Fig. 12.15. This gives kronecker deltas on the right given the physics normalization we are using in this chapter. The normalization will be changed in chapter 14.

diagram of Fig. 12.16 is often written in the simplified form shown in Fig. 14.7[15]

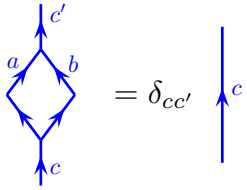


Fig. 12.17 A simplified version of the innner product in Fig. 12.16. This gives a kronecker delta on the right in the physics normalization we are using in this chapter. The normalization will be changed in chapter 14.

The fact that c must equal c' in Figs. 12.16 and 12.17 is consistent with the no-transmutation principle Fig. 12.12.

The principle of orthonormality of vertices implies the useful result that a loop, as shown in Fig. 12.18, is given the value of unity (This is jst Fig. 12.17 where we have set $c = c' = I$ and not drawn the identity line). At the risk of being repetitive we once again note that we will change this normalization in chapter 14 below and in later chapters, although it is correct for this section.

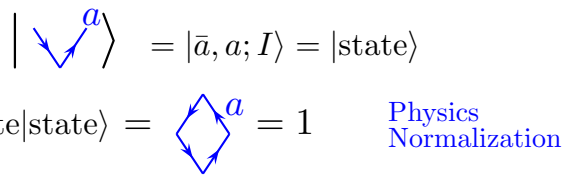


Fig. 12.18 The orthonormality of trees implies a particle loop gets a value of 1 if we are using physics normalization.

Since the vertex diagrams $|a, b; c\rangle$ from Fig. 12.14 form a *complete* set of states for the two particles, we can construct an identity operator for two strands as shown in Fig. 12.19.

We can derive a useful identity from Fig. 12.19 by applying projectors $|x\rangle\langle x|$ and $|y\rangle\langle y|$ to left and right strings respectively to obtain Fig. 14.8.

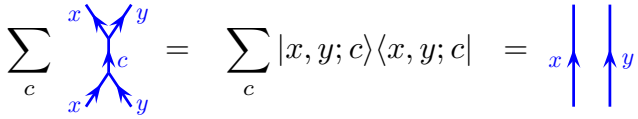
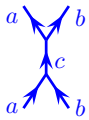


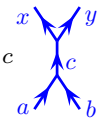
Fig. 12.20 Insertion of a complete set of states, with fixed quantum numbers x and y on both ends. This figure uses physics normalized diagrams. The normalization will be changed in chapter 14.

$$\sum_{a,b,c}$$ 

$$= \sum_{a,b,c} |a,b;c\rangle\langle a,b;c| =$$ 

Fig. 12.19 Insertion of a complete set of states. This figure uses physics normalized diagrams. The normalization wil be changed in chapter 14.

An arbitrary operator with two incoming and two outgoing lines can be written as in Fig. 12.21 where the coefficients $C_{(x,y),(a,b),c}$ are arbitrary (depending on what operator we want to define).

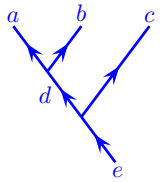
$$\sum_{a,b,x,y,c} C_{(x,y),(a,b),c}$$ 

$$= \sum_{a,b,c,x,y} C_{(x,y),(a,b),c} |x,y;c\rangle\langle a,b;c|$$

Fig. 12.21 An arbitrary operator with two incoming and two outcoming lines. The coefficients C are arbitrary.

12.2.3 Three Particles

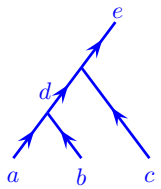
We can continue on and consider states of three particles. All the same principles apply here. As discussed in chapter 8, we can write an orthonormal complete set of states for three particles as a fusion tree[17,18] as in Fig. 12.22.



$$= |(a,b),c,d;e\rangle = |(a,b);d\rangle \otimes |d,c;e\rangle$$

Fig. 12.22 An orthonormal set of states for three particles can be described as a fusion tree. We have used two different notations on the right. The latter notation matches that of section 8.6.

If either $N_{ab}^{d} = 0$ or $N_{dc}^{e} = 0$ then the corresponding fusion is disallowed and the value of the diagram is zero[19]. The corresponding kets are obtained using the Hermitian conjugation rule of flipping the diagram and reversing arrows as shown in Fig. 12.23.



$$= \langle(a,b),c,d;e| = \langle(a,b);d| \otimes \langle d,c;e|$$

Fig. 12.23 The bras corresponding to the kets in Fig. 12.22.

The inner product of such states is given by stacking the bra on top of the ket as in Fig. 12.24.

[17] In cases where there are fusion multiplicities $N_{ab}^{d} > 1$ or $N_{dc}^{e} > 1$ then we must place an additional index μ or λ at the corresponding index. See for example, section 9.5.3.

[18] As mentioned in the above note 12, although we write this as a ket, it is really an operator, and to emphasize this we might write something like $|(a,b),c,d;e\rangle\langle e|$ instead.

[19] This is already implied by looking at the individual vertices and considering the rules of a single vertex as in Fig. 12.14.

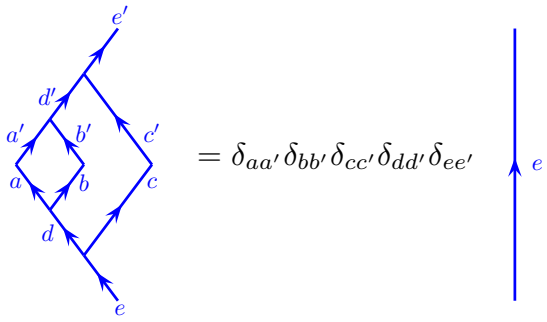


Fig. 12.24 The orthogonormality of tree states. This diagram uses physics normalization.

Note that in Fig. 12.24, one can focus one's attention on the left branches (with d, a, b, a', b', d'), which look exactly like Fig. 12.16 thus immediately obtaining $\delta_{aa'}\delta_{bb'}\delta_{dd'}$ and replacing the small diamond on the left branch with a single d-line. The remaining figure then looks exactly like the figure 12.16 and gives us the delta functions $\delta_{cc'}\delta_{ee'}$ as well. Thus the orthonormality of these tree states is not a seperate assumption but can be derived from the orthonormality of two particle states that we used in section 12.2.2.

The completeness of this set of states similarly can be expressed with diagrams as shown in Fig. 12.25.

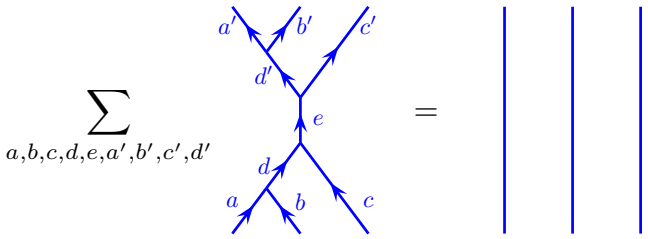
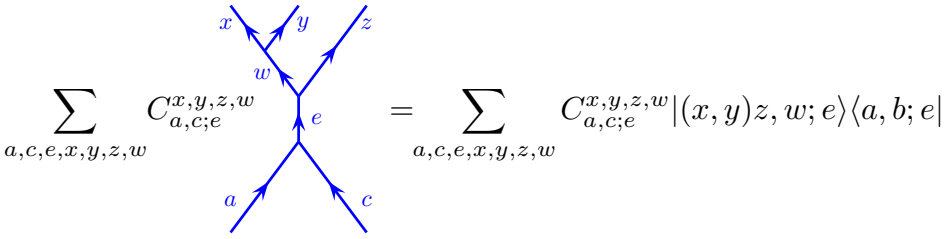


Fig. 12.25 The completeness of tree states for three particles. This diagram uses physics normalization.

Once again we can derive this completeness relation from what we know about the two-particle case. We can start in the very center of Fig. 12.25, considering the lines $d, c, e, d'c'$, and apply the completeness relation Fig. 12.20. This splits off the c-line to the right which, summed over its index gives a single unlabled line on the right as in Fig. 12.13. The remaining diagram on the left (with lines a, b, d, a', b' is of the form of Fig. 12.19 which summed over gives two unlabeled lines. Thus the completeness relation for three particles is not an independent assumption but follows from the completeness of the one and two particle cases.

One can use these basis states to build arbitrary operators with three particle states. Just as an example, in Fig. 12.26 we show the most general form of an operator that takes two particles as an input and gives three particles as an output.

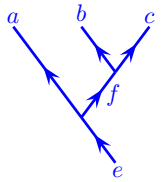


$$\sum_{a,c,e,x,y,z,w} C^{x,y,z,w}_{a,c;e} = \sum_{a,c,e,x,y,z,w} C^{x,y,z,w}_{a,c;e} |(x,y)z,w;e\rangle\langle a,b;e|$$

Fig. 12.26 An arbitrary operator with two incoming lines and three outgoing lines. The coefficients C are arbitary

12.2.4 F-Matrices Again

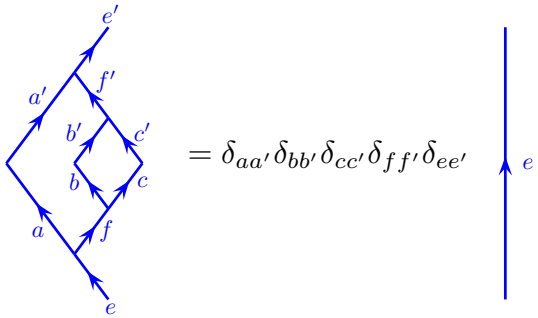
In defining our three particle states in Fig. 12.22 we have fused the two particles a and b on the left first to form d and then fused d with c to form e. (Our notation to indicate this is $|(a,b)c,d;e\rangle$). However, we could have chosen to fuse the particles in a different order to form a different tree as shown in Fig. 12.27. Here b and c fuse together to form f and then a and f fuse together to form e. We notate this state as $|a,(b,c),f;e\rangle$.



$$= |a,(b,c),f;e\rangle = |a,f;e\rangle \otimes |(b,c);f\rangle$$

Fig. 12.27 Another orthonormal set of states for three particles. Compare to Fig. 12.22.

The set of states defined by the fusion trees in Fig. 12.27 also form a perfectly good (but different) complete orthonormal basis of states for three particles. For example, we have the orthogonality relation shown in Fig. 12.28 (compare Fig. 12.24).



$$= \delta_{aa'}\delta_{bb'}\delta_{cc'}\delta_{ff'}\delta_{ee'}$$

Fig. 12.28 The orthogonality of tree states in a different basis. This diagram uses physics normalization.

As described in detail in chapter 9, if we draw trees with different branching structure, we are describing the same Hilbert space, but in a different basis — the basis change being given by a unitary F-matrix

transform as shown in Fig. 12.29 (See also discussion of Fig. 9.1).

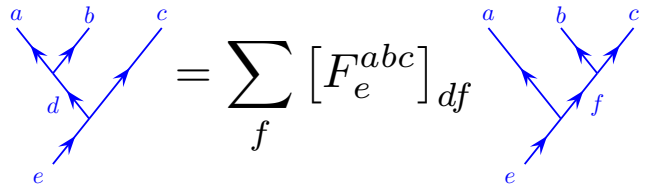


Fig. 12.29 The F-matrix. See chapter 9.

Similarly we have the relation between the Hermitian conjugate states as shown in Fig. 12.30. Note that because the F matrix is unitary in its

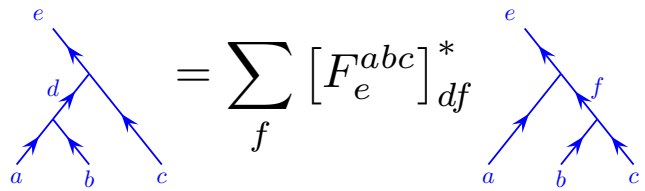


Fig. 12.30 F-matrix

two outside indices (d and f in Fig. 12.29) we have

$$[F_e^{abc}]^*_{df} = ([F_e^{abc}]^\dagger)_{fd} = ([F_e^{abc}]^{-1})_{fd}$$

12.2.5 More Particles

The principles we have developed for one-, two-, and three-particle states are easily extended to greater numbers of particles. Each shape of fusion tree defines a different orthonormal complete set of states. For example, with four particles, we might choose the tree shape shown in the left of Fig. 12.31, or we might choose the tree shape shown in the right of Fig. 12.31. Either one of these makes a perfectly good orthonormal basis for four particles — and these two bases are related to each other by F-matrices as discussed in chapter 9.

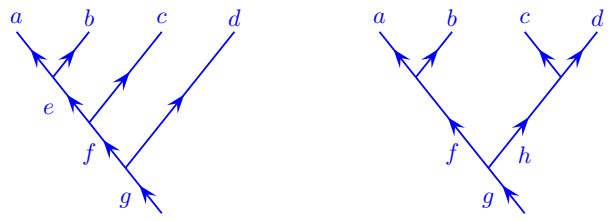


Fig. 12.31 Two (of five) possible bases for describing states of four particles. These bases are related to each other by F-moves (See Fig. 9.7). The shape of tree on the left is sometimes known as the "standard" basis.

The left-hand tree structure in Fig. 12.31, with all of the particles on

top branching from a single line going from top left to bottom right, is sometimes known as the "standard basis."

One can use F-moves to evaluate more complicated diagrams. An example of this is shown in Fig.12.32.

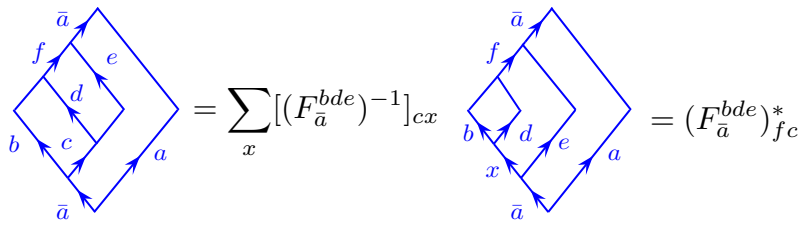


Fig. 12.32 The diagram on the left is evaluated by applying an F-move to the lower left part of the diagram. The resulting diagram is evaluated to a function δ_{xf} due to the orthonormality of tree diagrams. Finally we use the unitarity of F in the last step. Since this diagram starts and ends at the vacuum it evaluates to a scalar. This diagram is evaluated with physics normalization.

12.3 Causal Isotopy

Keeping with the idea of diagrams that are planar (no over- and under-crossings), we now consider how we may deform these diagrams. When we discussed the Kauffman bracket invariant we were allowed to freely deform any diagram as long as we did not cut any strands. This property is known as *isotopy invariance*[20]. Analogously, if a planar diagram retains the same value for any deformation that does not involve cutting strands or crossing them over each other, we say the theory has *full planar isotopy invariance*. Examples of this are shown in Fig. 12.33.

[20] In that case we had *regular isotopy* invariance meaning that we can deform knots freely in 3D as long as we treated the strands as ribbons. See sections 2.2.1 and 2.6.1.

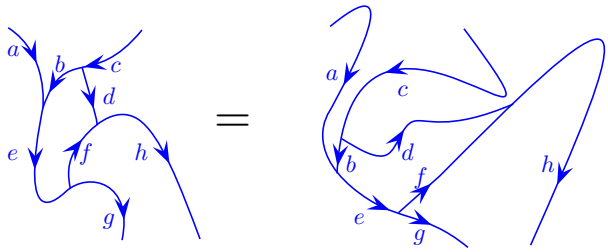


Fig. 12.33 For a theory with full planar isotopy invariance, these two diagrams should evaluate to the same result. Full planar isotopy invariance allows us to distort the diagram in any way as long as we do not cut any strands or cross lines through each other.

We need to ask how much topological invariance we should really expect from our physical theories. In the mathematical world of TQFTs and knot invariants, it is fine to assume that all directions are equivalent, and we can freely distort a line travelling in the x direction (horizontally) on the page to a line travelling in the t direction (vertically). However, in real physical systems, generically the time direction might need to

be treated differently from the space directions. In this section we will discuss topologically theories that allow deformation in space, but without allowing one to freely exchange the time and space directions. In particular some amount of causality might be demanded.

In chapter 16 we will consider a subset of theories which have a much higher level of topological invariance, known as *regular isotopy* invariance[21], which allows us to freely distort diagrams in either the space or time direction and further allow us to interchange the two.

In this chapter through chapter 15 we do not assume regular isotopy invariance (or full planar isotopy in the case of planar diagrams) but rather assume only what we call *causal isotopy*[22]. Here we allow deformation of space-time diagrams so long as we do not change the time-direction orientation of any lines or vertices. In other words, the path of a particle that is moving forward in time should not be distorted such that it is moving backwards in time (and vice-versa, a particle moving backwards should not be distorted so that it is moving forwards) — but other than this constraint, any smooth deformation is allowed. Two examples of deformations that are allowed under causal isotopy are shown in Fig. 12.34.

[21]The term "regular" implies that strands are treated as ribbons, but other than this caveat, all deformations without scissors are allowed. See section 2.6.1.

[22]This is not standard nomenclature.

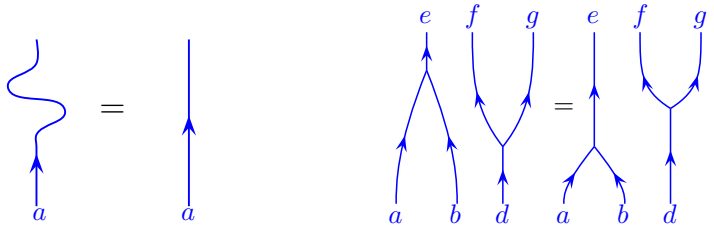


Fig. 12.34 Two examples of deformations that are allowed under causal isotopy. Deformations of the path are allowed as long as they do not require a particle to reverse directions in the time-like direction. In the left example, this deformation is allowed because in both cases the particle continues to move forward in the time direciton. In the right example, the temporal order of the vertices does not matter.

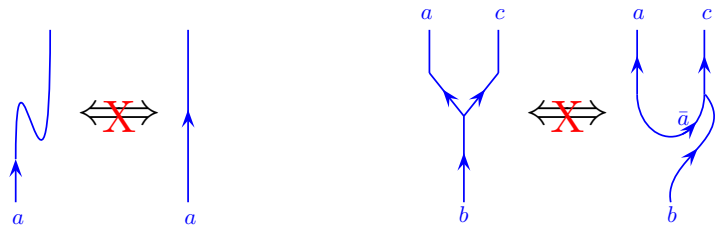


Fig. 12.35 Two examples of transformations that are not necessarily equalities under causal isotopy. In some special theories these transformations will be allowed, but generically they are not allowed. The diagrams on the far left are discussed in chapter 14. The diagram on the far right is discussed in sections 12.4.1 and 14.1.1.

Certain deformation of diagrams are not allowed by causal isotopy. Two examples of such disallowed deformations are given in Fig. 12.35. On the left of the figure we see a particle which turns around in time. This need not be the same as the particle moving straight in time as it

involves a particle creation event and a particle annihilation event. On the right of Fig. 12.35 a vertex is altered so instead of an a particle going out of the vertex, a $\bar{a}$ particle goes in. In this case we must have a a with $\bar{a}$ annihilation event in the far right diagram that does not exist in the simpler diagram where a and c directly fuse to b. Thus these two diagrams do not necessarily evaluate to the same result. (Although in some cases, such as in chapter 16, one may have a simple theory for which the tranformations showed in Fig. 12.35 are allowed, such theories are not generic.)

12.4 Summary of Planar Diagram Rules in Physics Normalization

With the principles we have now discussed we should be able to evaluate any planar diagram — taking a space-time process which starts and ends in the vacuum and turning it into an amplitude (i.e, a complex number). The same principles can be used to simplify operators such as Eq. 12.2.

Here are a summary of the important rules we have learned for diagram evaluation

(1) One is free to continuously deform a diagram consistent with causal isotopy as described in section 12.3. That is, particles must not change their direction in time due to the deformation.

(2) One is free to add or remove lines from a diagram if they are labeled with the identity or vacuum (I). See the example in Fig. 12.36.

(3) Reversing the arrow on a line turns a particle into its antiparticle (See Fig. 8.4).

(4) Regions must maintain their quantum number locally as in Fig. 12.12. In particular this means that a line must maintain its quantum number unless it fuses with another line, or splits.

(5) Splitting and fusion vertices are allowed for fusion multiplicities $N_{ab}^{c} > 0$ (See section 8.3). This includes particle creation and annihilation as a special case where a particle-antiparticle pair fuse to the vacuum or split from the vacuum (An example is shown in Fig. 12.36).

(6) Hermitian conjugation is given by reflection of a diagram around a horizontal line along with flipping the direction of arrows (See Fig. 12.4 or for example, Fig. 12.23)

(7) One can use F-moves to change the structure of fusion trees in order to simplify. For example, in Fig. 12.32.

(8) Once one reduces a diagram into tree structures that have the same branching in the upper and lower half (as on the right of Fig. 12.32) we can use the orthonormality of trees to complete the evaluation. In cases where the diagram starts and ends in the vacuum this reduces the diagram to a complex number (See, for

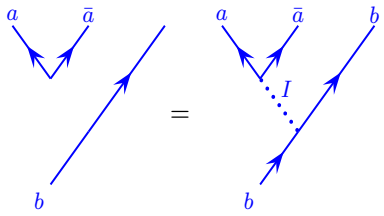


Fig. 12.36 One can always add or remove the identity (or vacuum) line to any diagram.

example, in Fig. 12.32). More generally operator diagrams can be reduced to simple forms analogous to Fig. 12.26.

With these principles (and given an F-matrix as input information – which will depend on the particular physical system we are considering) it is possible to fully evaluate any planar diagram, starting and ending in the vacuum, into a complex number. While there may be many strategies to use these rules to reduce a complicated diagram to a single complex number, the final result is independent of the order in which we apply the rules[23].

[23] This is guaranteed by the pentagon relation and the Mac Lane coherence theorem.

The mathematical structure we have defined thus far (our Hilbert space and F matrices) is known as a "unitary fusion category". There is more structure to be uncovered in further chapters that follows from what we have defined so far, and there are many special cases to be discussed. In addition note that here we have only described planar diagrams, so we have not yet described 2+1 dimensional theories — in order to describe these, we will have to include braiding rules for our diagrams we will add in chapter 13.

12.4.1 A Simple Example

As a simple example, let us try to evaluate the diagram shown on the far left of Fig. 12.38. We first work on a small part of the diagram as shown in Fig. 12.37 (Note that this is the same as the far right of Fig. 12.35).

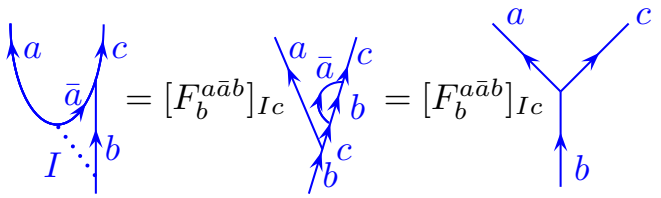


Fig. 12.37 To evaluate the diagram on the left, the vacuum line is inserted and an F-move is made. The bubble is then removed with Fig. 12.16. These diagrams use physics normalization. We will re-examine this diagram using a different normalization in section 14.1.1.

The result in Fig. 12.37 can also be reflected along the horizontal axis as in Fig. 12.4 to give the Hermitian conjugate diagram. Using both Fig. 12.37 and its reflection, we obtain the result given in Fig. 12.38.

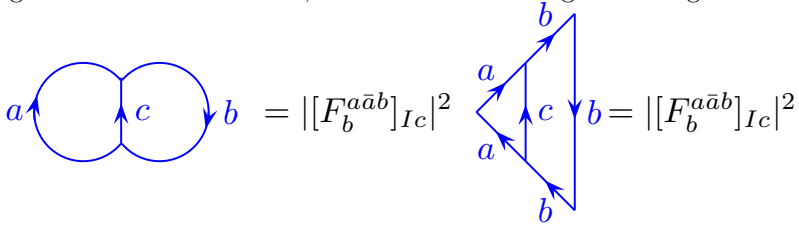


Fig. 12.38 The first step invokes Fig. 12.37 and its Hermitian conjugate. The figure on the right is a tree which evaluates to the identity so long as the fusion vertices are allowed and assuming physics normalizations.

12.5 Appendix: Higher Fusion Multiplicities

When we have a theory with higher fusion multiplicities (i.e., $N_{ab}^c > 1$ for at least one fusion channel), then the vertices must be given indices as well as lines having indices, and tree states are orthogonal in these indices as well. For example, we would need to modify Figs. 12.17 and 12.20. to the form shown in Figs. 12.39 and 12.40. See also the discussion of the F-matrix with higher fusion multiplicites in section 9.5.3.

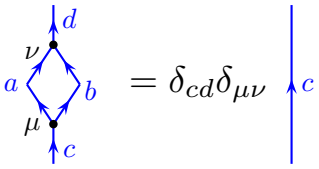


Fig. 12.39 The bubble diagram when there are fusion multiplicities. This diagram is a result of the orthonormality of tree diagrams. The variables at the vertices must match in order for the result to be nonzero. This diagram is drawn in the physics normalization. We will change the normalization in chapter 16.

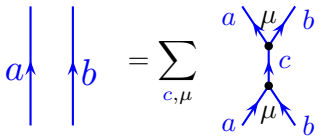


Fig. 12.40 Insertion of a complete set of states. When there are fusion multiplicities, these must be summed over as well $\mu \in N_{ab}^c$. This diagram is drawn in the physics normalization. We will change the normalization in chapter 16.

Chapter Summary

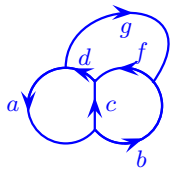
- THis is an item

Further Reading

This is some reading.

Exercises

Exercise 12.1 Evaluating diagrams with F-matrices
Evaluate the following diagram, writing the result in terms of F's.



Exercise 12.2 Locality Principle

Show that the locality principle (Fig. 12.12) is derivable from our other rules for evaluating diagrams, and is not therefore an independent assumption.

13

Braiding Diagrams[1]

In chapters 8, 9, and 12 we focused on planar diagrams. These diagrams can be thought of as describing the physics of objects that live in 1+1 dimension. More to the point, the nontrivial physics we discovered is really just a reflection of the nontrivial structure of the Hilbert spaces we are working with.

Here we extend our diagrammatic rules to the 2+1 dimensional world. In particular we want to describe what happens when we braid world lines. In chapter 10 we started to discuss braiding of identical particles and we continue that discussion here.

[1]This chapter continues the development of the diagrammatic algebra in some detail. Those who would like a brief and easier (albeit not as general) introduction to diagrammatic algebra should go straight to chapter 16.

13.1 Three Dimensional Diagrams

We begin by generalizing the concept of a diagram that we developed in chapters 9-16. The diagrams we want to consider now allow over- and under-crossings of lines as in Fig. 13.1. We will end up with a set of rules that are conceptually similar to the knot-invariants we discussed way back in chapter 2 — we take a picture of a generalized knot (like Fig. 13.1) and we reduce it to an output number. The generlization here is that the lines have labels $(a, b, c, \ldots)$ and lines can fuse with each other in addition to crossing over and under each other.

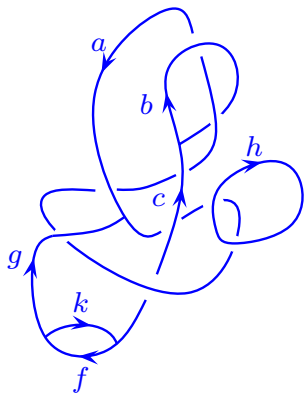


Fig. 13.1 A diagram with over- and under-crossings representing a process in 2+1 dimensions.

We should be somewhat cautious here that whereas when we considered the Kauffman bracket invariant, we had regular isotopy invariance — meaning that, treating strands as ribbons, any deformation of the diagram was allowed as long as we did not cut any strands. In contrast here (while we should still treat strands as ribbons) not all deformations are allowed. In general we will only have the same type of causal isotopy as described in section 12.3 (that is, we cannot freely deform a particle line going forward in time to one that goes backwards in time). Of course there do exist anyon theories with a higher level of isotopy invariance (regular isotopy), which we will discuss in chapter 16, but we should realize that these are not generic.

Our rules for evaluating diagrams with over- and under-crossings will be a consistent extension of a set of rules for evaluating planar diagrams[2]. Our next task is to consider how we handle over- and under-crossings. With this information, used in conjunction with the rules we have already developed for planar algebras, we will be able to evaluate any diagram in 2+1 dimensions.

[2]In mathematical language, the rules introduced in this chapter give additional structure to a unitary fusion category to make it a unitary braided fusion category, or unitary ribbon fusion category (these notions are equivalent.

13.2 Braiding Non-Identical Particles

We now turn to ask what happens if we exchange two different particle species, say a and b. We might be tempted to do something similar to Figs. 10.1 and 10.2 — that is we define a state with two particles in a given fusion channel then we exchange the two particles and determine the phase accumulated in this process. However, such a scheme cannot work in the case of non-indentical particles. The reason this fails is that when the two particles are not identical the initial and final states are fundamentally different and cannot be related to each other by just a phase — for example, the initial state for Fig. 13.2 has a to the left of b whereas the final state has a to the right of b.

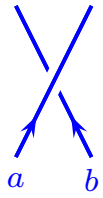


Fig. 13.2 One would ideally like a rule for exchanging any two particles. However, this will not generally be just a phase since the initial and final states are fundamentally differ from each other.

Nonetheless, the R-matrix can still be precisely defined even when we are braiding nonidentical particles. Diagrammatically we define the R matrix as shown in Fig. 13.3. On the right of this figure, the particles b and a come from c, with a going off to the right and b to the right. In the left of the figure, the two particles are moved away from each other, b to the right and a to the left, before they are braided around each other. The key here is that in both cases, the final state of the system has b on the left and a on the right, and the two particles fuse to a quantum number c, so that the two processes can be compared to each other and differ from each other only by a phase, which we define[3] to be R_c^{ab}.

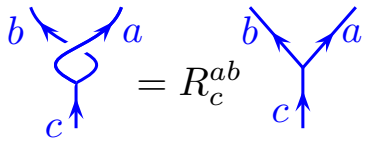


Fig. 13.3 Definition of the R-matrix. It is crucial that the final state of the system on both the left and right has b on the left and a on the right, and in both cases the two particles fuse to a. However, the left diagram includes an exchange of the two particles. The added exchange accumultates the phase R_c^{ab}.

Similarly we have the inverse braid

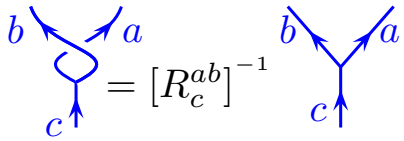


Fig. 13.4 The inverse exchange gives $[R_c^{ab}]^{-1}$.

It is important to note that (in a unitary theory) the R-matrix is always just a complex phase. Note that R_c^{ab} is not defined if a and b are not allowed to fuse to c (i.e, if $N_{ab}^c = 0$). Further, note that braiding anything with the identity (vacuum) particle should be trivial,

$$R_a^{Ia} = R_a^{aI} = 1$$

.

[3] The notation we use matches that of Bonderson [2007]. In this convention, the element R_c^{ab} is associated with the diagram having b going out the top left and a going out the top right. One might have thought that R_c^{ba} would have been a more natural notation. (Secretly, when no one is looking, I sometimes use the other convention. But in this book we will match the convention of the other references.)

A full braid (two exchanges in the same direction) of two particles a and b fusing to c is given by $R_c^{ab}R_c^{ba}$ as shown in Fig. 13.5. Note that in the representation on the far left of the figure, in both the initial and final configurations of particles, the b particle is to the left of the a particle meaning that we can understand this process as simply incurring an Aharonov-Bohm-like phase dependent on the fusion channel.

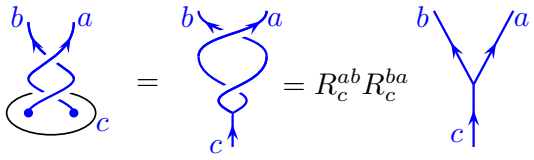


Fig. 13.5 A double exchange is a full braid (one particle wrapping fully around another.

If a particle a has trivial full-braiding with all other particles of a theory, ie., if $R_c^{ab}R_c^{ba} = 1$ for all a, c where $N_{ab}^c > 0$, then we call the particle type *transparent*. (The identity, or vacuum particle, is always transparent.)

Taken together with the F-matrices, the R-matrices allows us to calculate the physical result of any braid. The scheme is mostly analogous to the cases we discussed for braiding identical particles in chapter 10. If we want to exchange two particles we first use the F-matrices to put the system in a basis where those two particles have a known fusion channel. We can then directly apply the R matrix to describe the exchange.

In particular we can now give a general scheme for evaluating any crossing of the form shown in Fig. 13.2 which is shown in Fig. 13.6. Using this procedure any diagram with braiding can be reduced to a planar diagram which can then be evaluated using only the F-symbols.

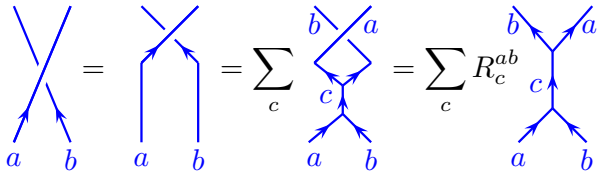


Fig. 13.6 A generic crossing can be reduced to a planar diagram using the R-matrix. In the second step a complete set of particles c is inserted as in Fig. 12.20. Note this figure uses physics normalization.

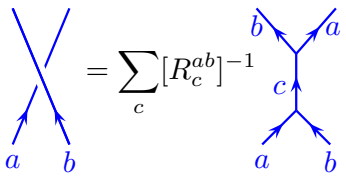


Fig. 13.7 The inverse crossing. This figure uses physics normalization.

Fig. 13.8 These moves, Reidermeister Type-II moves, are allowed in any anyon theory. See section 2.6.1.

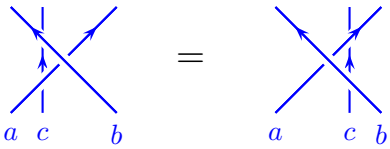


Fig. 13.9 This move, a Reidermeister Type-III move, is allowed in any anyon theory. See section 2.6.1.

Using the Hermitian conjugation principle (See Fig. 12.4) we can derive (see exercise 13.3)

$$[R_c^{ab}]^{-1} = [R_c^{ba}]^* \tag{13.1}$$

13.2.1 Summary of Rules for Evaluating any 2+1 D Diagram with Physics Normalization

The rules for evaluating any diagram in 2+1 dimensions (working with physics normalization of diagrams) are thus a very simple extension of the rules presented in section 12.4. We simply add two more rules

(1) We are allowed to use R-moves as in Fig. 13.3 and 13.4. In particular, this allows resolving of crossings by using Fig. 13.6 and 13.7.

(2) Once any diagram is reduced to a planar diagram, we can use the rules of section 12.4.

As with the case of planar diagram, there is some degree of deformation of diagrams (causal isotopy, see section 12.3) which is freely allowed. Here again the rules are similar: any deformation that does not involve cutting lines, or changing the time-direction of motion is allowed. Without introducing new assumptions, natural moves such as those shown in Fig. 13.8 and 13.9 can be derived (See exercise 13.5). These are nothing more than the Reidermeister Type II and III moves introduced in section 2.6.1.

13.3 The Hexagon

Using R-moves and F-moves any 2+1 D diagram (starting and ending in the vacuum) can be reduced to a complex amplitude. One might worry if the rules we have listed for evaluation of diagrams are self-consistent: i.e., does it matter in which order we apply the rules? Will we always obtain the same complex amplitude result? Indeed, given an F-matrix, only certain sets of (physically acceptable) R-matrices will

have the property that the diagrammatic rules give a unique final result. In fact, it is even possible that for a given set of F-matrices that satisfy the pentagon, there may not even exist a set of consistent R-matrices!

When we discussed planar diagrams in chapter 12, the pentagon equation guaranteed self-consistency of F-matrices. Now, given some F-matrices that satisfy the pentagon equations, the consistency equations for R-matrices are known as the hexagon equations and are shown diagrammatically in Fig. 13.10.

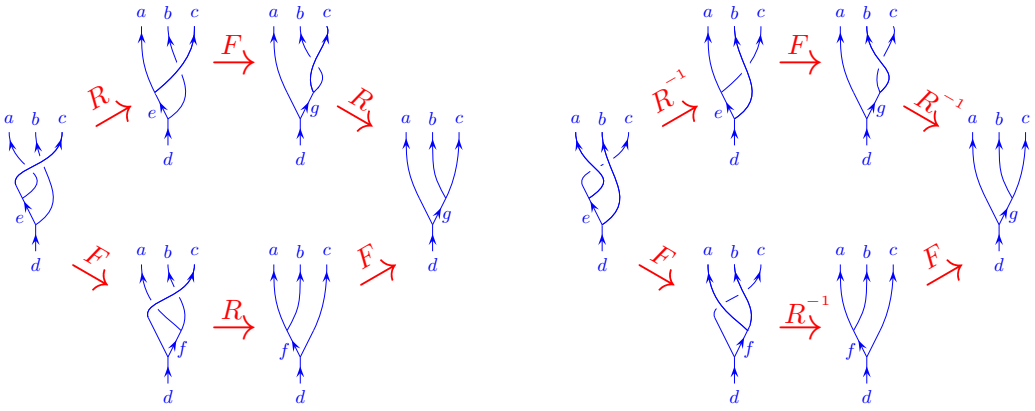


Fig. 13.10 The hexagon equations in graphical form.

In equations the hexagon conditions can be expressed as

$$R_e^{ca}[F_d^{acb}]_{eg}R_g^{cb} = \sum_f [F_d^{cab}]_{ef}R_d^{cf}[F_d^{abc}]_{fg} \qquad (13.2)$$

$$\left[R_e^{ca}\right]^{-1}\left[F_d^{acb}\right]_{eg}\left[R_g^{cb}\right]^{-1} = \sum_f [F_d^{cab}]_{ef}[R_d^{cf}]^{-1}[F_d^{abc}]_{fg} \qquad (13.3)$$

The top equation is the left diagram whereas the lower equation is the right diagram in Fig. 13.10. 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.

The structure we have now defined — a consistent set of (unitary) F and R-matrices satisfying the pentagon and hexagon equations, is known as a *unitary braided tensor category*[4] . All 2+1 D anyon theories must be of this form.

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.

With simple fusion rules, such as Fibonacci (as we saw in exercise 9.4) the fusion rules completely determine the F-matrices of the theory. Even with more complicated fusion rules, as we mentioned in section 9.3, there are only a finite number of possible solutions of the pentagon equation[5].

Once the F-matrices are fixed, there are only a finite number of pos-

[4] This is also sometimes known as a *unitary ribbon tensor category* due to the fact that Eq. 15.3 holds, which is always true for unitary theories with braidings. The unitary braided tensor category is also sometimes known as a *premodular* category.

[5] Solutions that can be obtained from other solutions by gauge transform are not counted as being different solutions.

sible solutions of the hexagon equations. For example, in the case of the Fibonacci theory, there are exactly two consistent solutions to the Hexagon equations, corresponding to the left and right handed types of Fibonacci anyons (See Eq. 10.2 and exercise 13.1).

13.4 R-matrix Odds and Ends

13.4.1 Appendix: Gauge Transforms and R

As in section 9.4 one can gauge make gauge transformations on the vertices of a theory. Given the transform shown in Fig. 9.8,

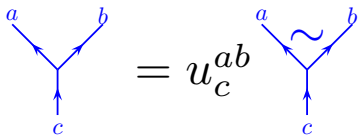


Fig. 13.11 We have the freedom to make a gauge transform of a vertex by multiplying by a phase u_c^{ab}. The tilde on the right notates that the vertex is in the tilde gauge.

the R matrix transforms as

$$\widetilde{R_c^{ab}} = \frac{u_c^{ba}}{u_c^{ab}} R_c^{ab} \tag{13.4}$$

Note that R_c^{aa} is gauge invariant, as is the full braid $R_c^{ab} R_c^{ba}$ in Fig. 13.5.

13.4.2 Product Theories

Given two anyon theories T and t, we can construct the product theory $T \times t$ as in section 8.5. If the theory T has consistent R matrices R_C^{AB} and the theory t has consistent R matrices R_c^{ab} ("consistent" here means there are F matrices that satisfy the pentagon relation and the F's and R's satisfy the hexagon relations), then the product theory has a consistent R matrices

$$R_{(C,c)}^{(A,a)(B,b)} = R_C^{AB}\, R_c^{ab}$$

Again, the point here is that in a product theory, the two consituent theories don't "see" each other at all.

13.4.3 Appendix: Higher fusion multiplicities

When we have a theory with higher fusion multiplicities (i.e., $N_{ab}^c > 1$ for at least one fusion channel), then the vertices must be given indices as well as lines having indices as in section 12.5. In this case the R-matrix carries vertex indices as well, and is a unitary matrix with respect to these indices.

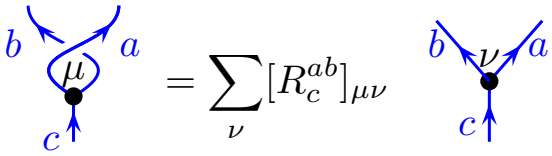


Fig. 13.12 Definition of R-matrix when there are higher fusion multiplicities. Here the vertices carry labels, and R is a unitary matrix with respect to these labels.

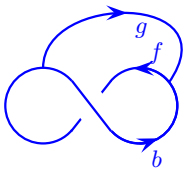
Under a gauge transformation, as in Fig. 9.10, the R matrix transforms as

$$\widetilde{[R_c^{ab}]}_{\mu'\nu'} = \sum_{\mu,\nu} ([u_c^{ab}]^{-1})_{\mu'\mu} [R_c^{ab}]_{\mu\nu} [u_c^{ba}]_{\nu\nu'} \tag{13.5}$$

Exercises

Exercise 13.1 Fibonacci Hexagon 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. 13.10. or the Fibonacci anyon theory, once the F matrix is fixed as in Eq. 9.3, 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. Confirm Eqs. 10.2 as one of the two solutions and show no other solutions exist.

Exercise 13.2 Evaluation of a Diagram
Consider the following diagram:



Evaluate this diagram in terms of R's and F's. Hint: First reduce the diagram to that shown in exercise 12.1.

Exercise 13.3 Symmetry of R
Use hermiticity to derive the property of R given in Eq. 13.1.

Exercise 13.4 Gauge transform of R and Hexagon
(a) Confirm the gauge transform Eq. 13.4.
(b) Show that a set of F matrices and R matrices satisfying the hexagon equations, Eq. 13.2 and 13.3 remains a solution after a gauge transformation. Remember that both R and F transform.

Exercise 13.5 Reidermeister Moves
(a) Use the R-matrix, and the completeness relationship, to derive the equivalence shown on the left of Fig. 13.8.
(b) How does the hexagon equation imply the equivalence shown in Fig. 13.13. Hint: This is very subtle, but is almost trivial.
(c) Use Fig. 13.13 to show the equality on the right of Fig. 13.8.

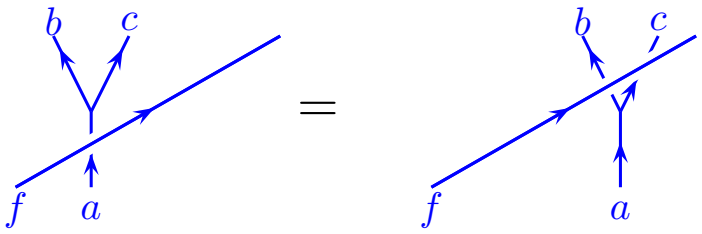


Fig. 13.13 This move is implied by the hexagon equation. (Similar with the straight line f going under the other two, and similar if the left-to-right slope of f is negative instead of positive.).

(d) Use the result of Fig. 13.13 along with completeness and the R-matrix to demonstrate Fig. 13.9.

This exercise shows that equalities like those shown in Fig. 13.8 and 13.9 are not indpendent assumptions but can be derived from the planar algebra and the definition of an R-matrix satisfying the hexagon.

14

Seeking Isotopy[1]

When we discussed knot invariants, like the Kauffman bracket invariant, we were allowed to deform a knot in arbitrary ways so long as we didn't cut any strands[2]. This is what we called isotopy invariance. We would very much like the diagramatic rules of our topological theories to obey isotopy invariance. However, as we discussed in section 12.3 we should really only expect invariance under a more limited set of moves which we called *causal isotopy*.

Fortunately, in many cases we can make some small changes to normalizations to remove some impediments to isotopy invariance. Let us first examine where the most obvious problem lies. For a nice topological theory (meaning one with full planar isotopy for a planar theory, or regular isotopy for a 2+1 dimensional theory) we would want to have the so-called zig-zag identity shown in Fig. 14.1 (which is not a property of theories having only causal isotopy invariance as mentioned in Fig. 12.35).

Unfortunately, a set of F matrices (even if they satisfy the pentagon self-consistency condition Eq. 9.7) does not generically satisfy this zig-zag identity Fig. 14.1. To see this, consider the manipulations shown in Fig. 14.2. With the physics normalization of diagrams we have been using, the zig-zag identity does not hold.

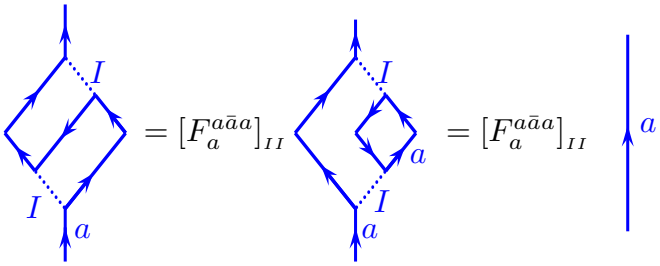


Fig. 14.2 Straightening a zig-zag wiggle incurs a factor of F using physics normalization of diagrams. The left of this diagram is the same as the left of Fig. 14.1. In the first step we use an F-move on the lower part of the diagram. We then use orthogonality of the tree to remove the small a bubble. This part of the diagram is just Fig. 12.18. Thus this small a bubble can be removed. We conclude that with the physics normalization we cannot satisfy the zig-zag identity Fig. 14.1.

14.1 Isotopy Normalization of Diagrams

To fix the zig-zag problem, we take a cue from the Kauffman bracket invariant and change our definition of diagrams just by a small bit. In

[1] This chapter continues the development of the diagrammatic algebra in some detail. Those who would like a brief and easier (albeit not as general) introduction to diagrammatic algebra should go straight to chapter 16.

[2] Meaning regular isotopy — i.e., we should treat strings as ribbons. See section 2.2.1.

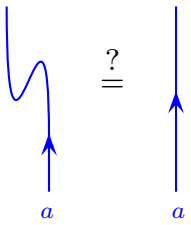


Fig. 14.1 A topological theory with full isotopy invariance should have this "zig-zag" identity. However, generically a set of F matrices will not satisfy this equality (See Fig. 14.2). We can often repair this problem by changing the normalization of kets.

particular, let a simple loop of particle a, as shown in Fig. 14.3, be given a value which we call d_a. This is different from our prior definition where we set the loop value to one as in Fig. 12.18. The change here only means that we will be working with unnormalized bras and kets. We will call this normalization "isotopy normalization".

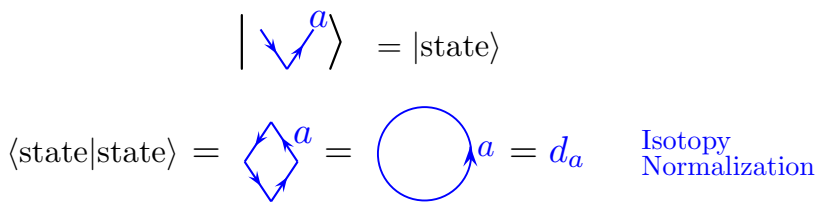


Fig. 14.3 Using a new normalization (which we call "isotopy normalization") of bras and kets. Compare to Fig. 12.18.

We should not worry about working with unnormalized bras and kets — we are allowed to do this in quantum mechanics. The price for using unnormalized states is that expecations of operators are now given by

$$\langle \hat{O} \rangle = \frac{\langle \psi | \hat{O} | \psi \rangle}{\langle \psi | \psi \rangle}$$

instead of the usual expression for normalized states which just has the numerator. Note that clearly for the identity particle $d_I = 1$ since we should be able to add and remove and deform vacuum lines freely[3].

[3]Evaluation of an empty diagram also gives unity, since one can add or remove vacuum lines freely, we can think of the empty diagram as being equivalent to any number of loops of the vacuum I.

Henceforth, we will use isotopy normalization!

With this new normalization, we can recalculate the value of a zig-zag wiggle analogous to that of Fig. 14.1.

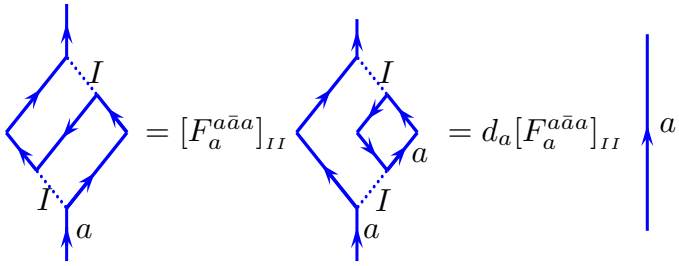


Fig. 14.4 With the new isotopy invariant normalization of diagrams, straightening a zig-zag wiggle incurs a factor of $d_a[F_a^{a\bar{a}a}]_{II}$. We will choose the value of d_a so as to make this factor unity.

If we can arrange that the prefactor $d_a[F_a^{a\bar{a}a}]_{II}$ is unity then straightening a zig-zag such as that in Fig. 14.1 will be an allowed transformation. In the simplest theories, we can simply choose d_a such that this product is unity. However, this is not always going to be possible to do (this will be discussed in more detail in section 18.1.2). Instead, what we will always be able to do (which is *almost* what we would like!) is to arrange that $[F_a^{a\bar{a}a}]_{II}$ is real[4] and we choose d_a such that

[4]This is arranged by gauge transform. See section 18.1.2.

$$d_a = \frac{\epsilon_a}{[F_a^{a\bar{a}a}]_{II}} \tag{14.1}$$

where

$$\epsilon_a = \pm 1 \tag{14.2}$$

which we will discuss in more detail in section 18.1.2 below[5]. Given this choice, the product $d_a[F_a^{a\bar{a}a}]_{II}$ in Fig. 14.4 is $\epsilon_a = \pm 1$ and the the zig-zag identity generally becomes modified to that shown in Fig. 14.5.

[5] I like to call ϵ_a the "wiggle phase" in absence of a better name.

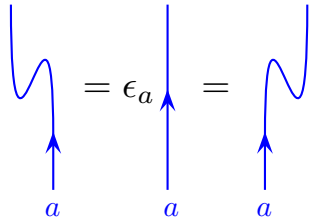


Fig. 14.5 The modified zig-zag identity. Here ϵ_a is always arranged to be $+1$ or -1. The equality on the left is from Fig. 14.4. The equality on the right follows from Hermitian conjugation of the equality on the left (turning the diagrams upside down and reversing the arrows).

Thus by changing the normalization of a loop in Fig. 14.3 and choosing the value d_a of this loop appropriately, we arrange such that zig-zag lines as in Fig. 14.5 can be freely straightened out, up to a possible sign. In the simplest theories, $\epsilon_a = +1$ for all particle types and zig-zag lines can be straightened out freely without accumulating any sign. However, this will not always be the case.

Thus we have defined a new normalization of the loop Fig. 14.3 given by the choice of Eq. 14.1. As we will see in section 17.1, the normalization constant d_a will turn out (up to a possible sign[6]) to be the same quantum dimension d_a that we found in Eq. 8.9 from the Hilbert space dimension of fusing anyons together!

Having changed the normalization of our kets, for consistency we need to change the normalization of fusions and splittings vertices as well. Thus we define new normalization of vertices as shown in Fig. 14.6[7]

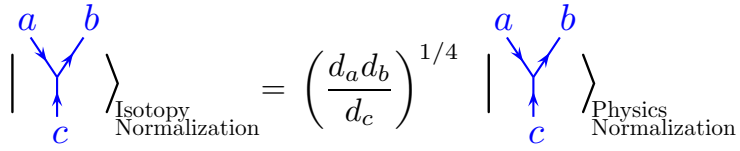


Fig. 14.6 New "isotopy" normalization for vertices[7]. Note that this is consistent with Fig. 12.18 by setting $c = I$ with $a = b$ (and note that $d_I = 1$.).

With this new normalization, the orthonormality of trees is now different from what we previously assumed. For example, Fig. 12.32 should now have a factor of $\sqrt{d_a d_b d_e d_d}$ on the right hand side.

Similarly our bubble diagram Fig. 12.17 and our completeness diagram Fig. 12.20 need to be modified as shown in Fig. 14.7 and Fig. 14.8[8].

[6] Note that our conventions allow d_a to be positive or negative. We will discuss potential issues with $d_a < 0$ in section 18.1.2 below. Some references, including Bonderson [2007] and Kitaev [2006] define d_a to be the absolute value of this quantity so it is the same as d_a. We have already seen examples where $d_a < 0$ (see exercise 2.2).

[7] One might be worried how one handles the fractional power when one of the d's is negative. In fact we will only need to consider cases where the factor inside the brackets ends up positive. (See Eq. 14.4 below). This is always possible since we can always choose to put the minus sign on ϵ rather than d.

[8] Once again if $N_{ab}^c > 1$ there are additional indices at the vertices and these must match as well. See section 12.5.

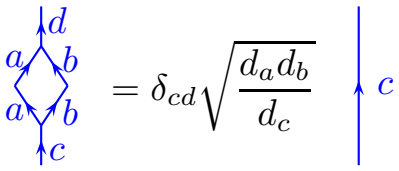


Fig. 14.7 Bubble diagram with isotopy invariant normalization of diagrams. See Eq. 14.5 for how to interpret the square root in cases where $d < 0$.

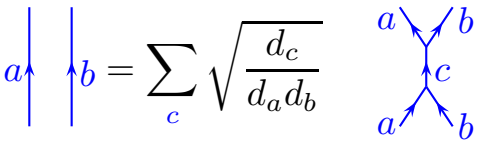


Fig. 14.8 Insertion of a complete set of states with isotopy invariant normalization of diagrams. See Eq. 14.5 for how to interpret the square root in cases where $d < 0$.

A crucial point is that the F-matrix does not need any alteration when we switch from physics normalization to isotopy normalization[9]! One can check that in changing normalizations both sides of Fig. 9.1 (equivalently Fig. 12.29) are multiplied by the same factor of $(d_a d_b d_c / d_e)^{1/4}$.

With this isotopy invariant normalization the rules for evaluating planar diagrams are exactly the same as those described in section 12.4 except that loops are now normalized with the quantum dimension as in Fig. 14.3 and our orthonormality relationships (Fig. 12.20 and Fig. 12.17) are altered to those shown in Fig. 14.8 and Fig. 14.7.

The same R-matrix rules can be applied to diagrams with over- and under-crossings as in chapter 13. The use of the R-matrix is unchanged. Be warned, however, that in Fig. 13.6 and 13.7 we have used the completeness relationship Fig. 12.20 which now needs to be modified to Fig. 14.8, so that when we evaluate crossings we now obtain, for example, Fig. 14.9.

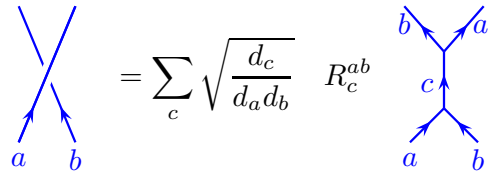


Fig. 14.9 Resolving a crossing with isotopy normalization. Compare to Fig. 13.6. See Eq. 14.5 for how to interpret the square root in cases where $d < 0$.

14.1.1 Futher Possible Impediments to Full Isotopy Invariance

With this new isotopy invariant normalization we allow straightening of wiggly lines (i.e., the zig-zag identity is obeyed) as in Fig. 14.1 up to a possible sign, which we will discuss further in section 18.1.2. However, even neglecting this sign, we emphasize this does not guarantee full planar isotopy invariance, that is that we can deform lines in any way we like in the plane. For example, the right hand side of Fig. 12.35

[9] Indeed, the reason why we changed the value of all vertices, as in Fig. 14.6, and not just rescale the vertex corresponding to a simple loop as in Fig. 14.3, is in order to keep F from changing.

cannot generically be turned into an equality. In Fig. 14.10 and 14.11 we give similar examples (Recapitulating the calculation in Fig. 12.37 but now using isotopy normalization) of turning-up transformations that generically incur nontrivial factors.

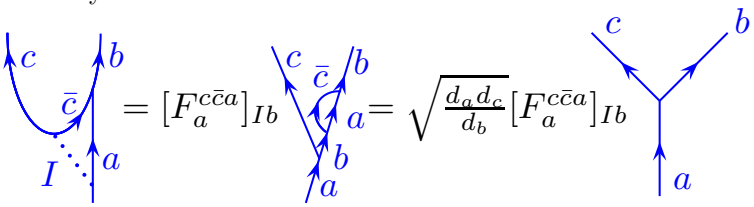


Fig. 14.10 To evaluate the diagram on the left, the vacuum line is inserted and an F-move is made. The bubble is then removed with Fig. 14.7. Note that if we were to use the physics normalization, the prefactor of $\sqrt{d_a d_c/d_b}$ would be absent (See Fig. 12.37). Generally we should not expect that the prefactors of d's and F obtained on the right should cancel each other. In chapter 16 we focus on precisely the theories where this does turn out to be unity as is required for full isotopy invariance. More generally, as we will discuss in section 14.7.1, the transform from left to right in this figure is unitary, meaning the resulting factor on the right $\sqrt{d_a d_c/d_b}[F_a^{c\bar{c}a}]_{Ib}]$ is just a magnitude one complex phase.

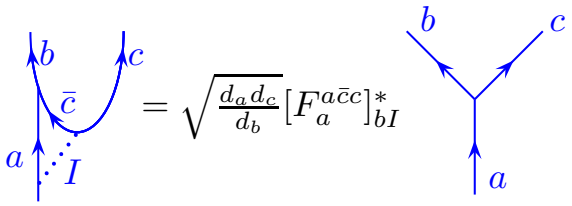


Fig. 14.11 The mirror image of Fig. 14.10. Here we use the fact that F is Hermitian, so $F^{-1} = [F^*]^T$.

Thus it seems that our most general theory with causal isotopy invariance cannot achieve full planar isotopy invariance. Perhaps this is not surprising. Even if we can deform space-time world lines into each other, we might still expect that there would be some minor difference between a process on the far left and far right of Fig. 14.10: On the far left c and $\bar{c}$ are produced from the vacuum then $\bar{c}$ and a come together to form b whereas on the far right, a simply turns into c and b. Fortunately, many topological theories are not this complicated: as we will see in chapter 16, there are many theories where one does have full planar isotopy invariance, and the prefactor incurred in the process shown in Fig. 14.10 turns out to be unity.

14.2 Gauge Choice and Frobenius-Schur Indicator

Let us now return to the zig-zag in Fig. 14.1 and our choice of the quantity d_a in Eq. 14.1. First, we claimed that we can always arrange to have $[F_a^{a\bar{a}a}]_{II}$ be real. With a gauge choice, we can fix the phase of $[F_a^{a\bar{a}a}]_{II}$ any way we like, at least for cases where $a \neq \bar{a}$. Let us see how

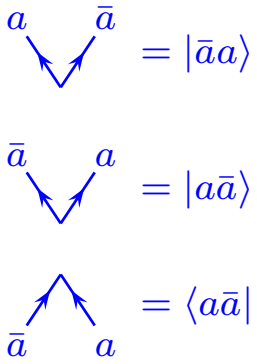


Fig. 14.12 The vertex $|a\bar{a}\rangle$ (top) and the vertex $|\bar{a}a\rangle$ (middle) can be assigned different phases as a gauge choice (See section 9.4). The bottom figure here is the Hermitian conjugate of the middle and must have the conjugate phase choice. In Fig. 14.4 the leftmost figure includes $|a\bar{a}\rangle$ and $\langle\bar{a}a|$, whereas the phases cancel in the loop formed in the middle picture of Fig. 14.4 which is formed from $|a\bar{a}\rangle$ (middle here) and $\langle a\bar{a}|$ (bottom here). Thus choosing gauges we can choose any phase for $[F_a^{a\bar{a}a}]_{II}$ unless $a = \bar{a}$ (See section 9.4 for discussion of the effects of gauge transform on F)..

[10]For particles which are not self-dual there are severeal different definitions of what people call the Frobenius-Schur indicator. Some references just define it to be zero for such particles. Other references define it to be $\epsilon_a \text{sign}[d_a]$. To avoid confusion we will not use the phrase Frobenius-Schur in the context of non-self-dual particles.

[11]Theories with all positve Frobenius-Schur indicators are sometimes called *unimodal* or *unimodular*.

this can be done. On the far left of Fig. 14.4 we have a vertex $|\bar{a}a\rangle$ as well as a vertex which we write as $\langle a\bar{a}|$ (compare to Fig. 14.12). Note that, at least when $a \neq \bar{a}$ these two vertices are *not* Hermitian conjugates of each other (recall that when Hermitian conjugating a diagram arrows get reversed as well as reflecting the diagram). By making separate gauge transforms on these two states, these kets can be redefined by an arbitrary phase as discussed in section 9.4, and this phase then ends up in $[F_a^{a\bar{a}a}]_{II}$ (See the transformation in Eq.). Thus by a gauge choice we can choose any phase for $[F_a^{a\bar{a}a}]_{II}$, as long as $a \neq \bar{a}$. It is often convenient to choose $[F_a^{a\bar{a}a}]_{II}$ to be real and positive so that we can have d_a positive and $\epsilon_a = +1$ postive as well. It is nice to have $d_a > 0$ because we want our states in Fig. 14.3 have positive norm, which is appropriate for quantum mechanics (although there may be cases where we choose d_a negative instead. See Eq. 14.4 below). Similarly it is nice to have $\epsilon_a = +1$ since this means we can straighten wiggles as in Fig. 14.5 without incurring signs.

However, if $a = \bar{a}$, it is not possible to change $[F_a^{aaa}]_{II}$ by gauge transform. In this case the kets $|a\bar{a}\rangle$ and $|\bar{a}a\rangle$ are equal and we do not have the freedom to gauge transform them separately. It is easy to show that when $a = \bar{a}$, the factor of $[F_a^{aaa}]_{II}$ must be real (See appendix 14.6 for a three line proof). The sign of $[F_a^{aaa}]_{II}$ is then a gauge invariant quantity, known as the Frobenius-Schur indicator[10]

$$\kappa_a = \text{sign}[F_a^{aaa}]_{II} = \epsilon_a \,\text{sign}[d_a] \quad . \tag{14.3}$$

If the Frobenius-Schur indicator is positive for all the self-dual particles in a theory, then we can set $\epsilon_a = +1$ for all particles and we can also have d_a positive for all particles. This means that we can both have a positively normed inner product, and we can freely straighten out wiggles as in Fig. 14.5 without incurring any minus signs. Theories of this type are fairly simple to work with[11].

However, when the Frobenius-Schur indicator of a self-dual particle a is negative, we can choose whether to put the minus sign in Eq. 14.1 onto d_a or to put it onto ϵ_a. Each of these has its disadvantages. On the one hand, having negative d_a means we are working with a non-positive-definite inner product. On the other hand, putting the minus sign on ϵ_a means we have to keep track of all the wiggles in our diagram as in Fig. 14.5. Either one of these approaches ends up giving complications in keeping track of signs. In appendix 14.4 we show that it is not so strange to have these sign complications, and such signs occur even for conventional spin 1/2 particles!

An alternate approach used by many mathematicians (Bakalov and Kirillov [2001]) is to create a fictitious degree of freedom and (even when $a = \bar{a}$) treat a and $\bar{a}$ as different objects. Effectively similar approaches are used by Lin and Levin [2014], as well as Kitaev [2006] and Bonderson [2007]. This scheme has advantages in that it appears fully isotopy invariant, but it comes at the cost of not fixing a gauge. We discuss these scheme, and its pro's and con's in section **??**.

14.2.1 Simple Bookkeeping Scheme

Here we will describe an extremely useful and faily general bookkeeping scheme for handling the Frobenius-Schur signs.[12]

If one chooses to have negative values of d_a it is important to arrange that

$$\text{sign}[d_a]\,\text{sign}[d_b] = \text{sign}[d_c] \qquad \text{when} \qquad N_{ab}^{c} > 0 \tag{14.4}$$

This condition assures that that factor $d_a d_b/d_c$ in Fig. 14.6 is positive, so that for example, the square roots in Figs. 14.7, and 14.8 have positive arguments. This nonetheless, leaves an ambiguity as to whether we should take the positive or negative square root. The convention we will use below is that

$$\sqrt{\frac{d_a d_b}{d_c}} = \begin{cases} \text{negative} & d_a < 0 \text{ and } d_b < 0 \\ \text{positive} & \text{otherwise} \end{cases} \tag{14.5}$$

Since we are able to make gauge choices for all non-self-dual particles, it turns out that it is almost always possible satisfy Eq. 14.4 and at the same time choose all $\epsilon_a = +1$ which makes diagrammatics particularly easy[13] since then one can freely straighten wiggles as in Fig. 14.5.

Thus we will choose $\epsilon_a = +1$ for all particles yet we may generally allow some $d_a < 0$ (this is only necessary if there exists at least one self-dual particle with negative Frobenious-Schur indicator). It may sound problematic to have some $d_a < 0$ since the negative normed state in Fig. 14.3 seem like they would violate the principles of unitary in quantum mechnanics. However, with a small reinterpretation of the meaning of our inner product, we can reinterpret our diagrammatics as representing a well behaved unitary theory.

Our reinterepretation of this diagrammatic algebra is quite simple. We evaluate diagrams using the rules given in section 14.1. That is, we use the rules from section 12.4 except that loops are now normalized with the quantum dimension as in Fig. 14.3 and our orthonormality relationships (Fig. 12.20 and Fig. 12.17) are altered to those shown in Fig. 14.8 and Fig. 14.7 (noting the choice of sign in square roots given by Eq. 14.5.) If there are over- and under-crossings, these can be evaluated using the R-matrix as in Fig. 14.9. Crucially, since we have set all $\epsilon_a = 1$, zig-zag wiggles like Fig. 14.5 can be freely straightened out (although recall that diagrams like 14.10 may not be freely straightened in general).

In the case where there are some $d_a < 0$, we call the result of this evaluation the *non-unitary* evaluation of the diagram as it corresponds to the non-unitary inner product. However, we now insert one additional rule into our list

(0) Before evaluating a diagram, count the number of negative-d caps, and call it n. After fully evaluating the diagram multiply the final result by $(-1)^n$.

Here a negative-d cap occurs when we go forward in time and two particles with $d < 0$ come together to annihilate or form a particle having $d > 0$. (See examples in Figs. 14.13 and 19.4). Another way of counting

[12] This scheme was constructed by Joost Slingerland and myself, but is so far unpublished.

[13] The simplest known, modular anyon theory for which one cannot both satisfy Eq. 14.4 and choose all $\epsilon_a = +1$ is a very complicated theory with 64 different fields. A simpler example of a planar diagram algebra where one cannot choose all $\epsilon_a = +1$ is discussed by Lin and Levin [2014].

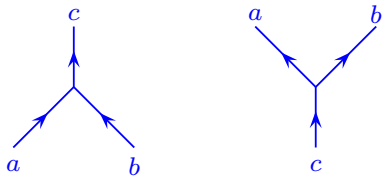


Fig. 14.13 With time going vertical, the left diagram is a neagtive-d cap if and only if $d_a < 0$ and $d_b < 0$. (The directions of the arrows do not matter, and if the particles are self-dual we do not draw arrows). The right diagram is never a negative-d cap.

Fig. 14.14 With time going vertical, the left diagram is a negative-d cap if and only if $d_a < 0$. The right diagram is never a negative-d-cap. We can think of these diagrams as being the same as the diagrams in Fig.14.13 with c being the identity. The directions of the arrows do not matter.

the negative-d caps is to imagine erasing all lines in the diagram which have $d > 0$. This leaves only a set of closed loops (due to Eq. 14.4). We then just need to count caps in this set of closed loop of the form shown in the left of Fig. 19.4.

With these new rules, we are now describing a unitary positive-normed quantum theory — we call this evaluation of a diagram, including rule 0, the *unitary* evaluation of the diagram. To understand the intuition behind these rules, consider a self-dual particle with negative Frobenius-Schur indicator. For such particles a wiggle like in Fig. 14.5 is supposed to incur a minus sign — however, in our scheme we have set $\epsilon = +1$ and instead made d negative. Since $\epsilon = +1$ in the diagrammatic algebra, there is no sign associated with straightening a wiggle. However, the wiggle in Fig. 14.5 has a negative-d cap, so in the final evaluation of the diagram (applying rule 0) we correctly obtain the required minus sign.

As a simple example, consider the evaluation of a single loop as in Fig. 14.3 where $d_a < 0$. Before evaluating the loop we count that there is a single negative-d cap on the top of the loop (as in the left of Fig. 19.4). We evaluate the diagram with the rules of section 16.1.2, to obtain $d_a < 0$ as the nonunitary evaluation. However, applying rule (0) this quantity is then multiplied by -1, giving the final result for the quantum dimension $-d_a = |d_a| > 0$. This is the result of the unitary evaluation of the diagram, and it is positive as we would hope for a positive definite inner product for a diagram that can be written as $\langle \text{state}|\text{state}\rangle$ (See Fig. 14.3.)

As a second example, consider the diagram Fig. 14.15, and let us assume that $d_a, d_b < 0$ and $d_c > 0$. The (nonunitary) evaluation of the diagram (without rule 0) gives $-d_c\sqrt{d_a d_b/d_c}$, the square root coming from Fig. 14.7 and the sign from the rule Eq. 14.5 of how to handle square roots with negative d's. However, applying rule 0, there is a single negative-d cap (from the vertex with a and b coming in from the bottom, and c going out the top), and hence the unitary evaluation of this diagram is $+d_c\sqrt{d_a d_b/d_c}$. Note that this is postive as it should be for a diagram that can be written as $\langle \text{state}|\text{state}\rangle$ analogous to Fig. 14.3.

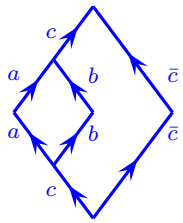


Fig. 14.15 An example of a diagram which should have a postive definite value since it can be written as $\langle \text{state}|\text{state}\rangle$.

As a third example, consider the same diagram Fig. 14.15 but consider the case where $d_a, d_c < 0$ and $d_b > 0$. Here the nonunitary evaluation gives $d_c\sqrt{d_a d_b/d_c}$, but applying rule 0, with a single negative-d cap (the top of the c loop) we obtain a final result of the unitary evaluation given by $-d_c\sqrt{d_a d_b/d_c}$. Note that this is also positive as it should be.

The situation described in this section — having a theory which allows straightening of wiggles, but has negative d_a, which can then be interpreted as a unitary theory — is quite common. There are many topological theories of this type — including the semion theory $SU(2)_1$ and more generally theories like $SU(2)_k$.

14.3 What have we achieved?

One of our original hopes for defining a TQFT, way back in chapter 7, was some prescription that would turn a labeled knot or link dia-

gram into a complex amplitude (see Fig. 7.1) where the result would be unchanged by any smooth deformation of space-time (treating the strands of the knots as ribbons, i.e., we are allowed regular isotopy of the diagram). At least to the extent that we can choose $\epsilon_a = +1$ for all particles, we have actually now achieved this goal![14]. The cases where we fail to have full planar isotopy invariance (in section 14.1.1) involve fusion vertices, and as long as our diagrams do not have such fusion vertices any deformation of the diagram is allowed[15].

Note that in chapter 7 when we were defining a TQFT we wanted to more generally have a prescription for turning a knot or link *embedded in an arbitrary closed manifold* into a complex number output. This generalization will indeed be possible, and we will return to this issue in chapter ***. However, for now we note that our scheme gives unity for an empty diagram (which we can think of as any number of loops of the identity particle with $d_I = 1$) so our diagrammatic evaluation corresponds to

$$\text{diagram} = \frac{Z(S^3 \text{ with labeled link embedded})}{Z(S^3)} \qquad (14.6)$$

$$= Z(S^2 \times S^1 \text{ with labeled link embedded})$$

with the caveat that the link does not go around the handle of the S^1 in the latter case.

[14] It is the *non-unitary* evaluation of the diagram that has full isotopy invariance. Once we apply rule 0, we incur minus signs for straightening a wiggle again, although we then obtain a unitary theory. The schme for handling Frobenius-Schur indicators discussed in section **??** does not have this problem and gives us completely isotopy invariant link invariants, but at the cost of not having a fixed gauge.

[15] It is interesting that in order to *evaluate* our knot or link diagrams we may need to turn these knotted lines into into fusing lines, such as in Fig. 14.9, which may then not be invariant under some smooth deformations as in section 14.1.1.

14.4 Appendix: Spin 1/2 Analogy

It may seem a bit odd that wiggling a space-time line (as in Fig. 14.5) can incur a minus sign. While this physics might appear a bit unfamiliar it turns out that there is a familiar analog in angular momentum addition — where the particle types (the labels a, b, c etc) correspond to the eigenvalue of total angular momentum squared J^2.

Consider three spin-1/2 particles which all taken together are in an eigenstate of $J = 1/2$. We can describe the possible states of the system with fusion trees as in Fig. 14.16 (see also Fig. 9.1)— in this case where a, b, c and e are all labeled with $J = 1/2$. In Fig. 14.16 we can (on the left of the figure) consider either the fusion of the left-most two particles to some angular momentum $d = 0$ (meaning a singlet) or $d = 1$ (meaning a triplet), or we can (on the right of the figure) consider fusion of the right-most two particles to either $f = 0$ or $f = 1$. The F-matrix that relates these two descriptions of the same space is given by $[F^{\frac{1}{2}\frac{1}{2}\frac{1}{2}}_{\frac{1}{2}}]_{df}$ which is often known as a $6j$ symbol in the theory of angular momentum addition. The analogy of negative Frobenius-Schur indicator here is the fact that $[F^{\frac{1}{2}\frac{1}{2}\frac{1}{2}}_{\frac{1}{2}}]_{00}$ is negative.

Let us try to see how this happens more explicitly. Given that the total spin is 1/2 we can focus on the case where the total z-component of angular momentum is $J_z = 1/2$ as well. The state where the leftmost

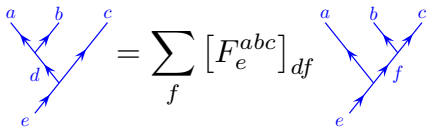


Fig. 14.16 The F-move.

two particles fuse to the identity (or singlet $J = d = 0$) can then be written explicitly as

$$|\psi\rangle = \frac{1}{\sqrt{2}} \left(|\uparrow_1\downarrow_2\rangle - |\downarrow_1\uparrow_2\rangle\right) \otimes |\uparrow_3\rangle \tag{14.7}$$

where the subscripts are the particle labels given in left to right order. This wavefunction is precisely analogous to the lower half (the "ket") of the far left hand picture in Fig. 14.2.

On the other hand, we could use a basis where we instead fuse the rightmost two particles together first, as in the righthand side of Fig. 14.16. We can write the state where the right two fuse to $J = f = 0$ analgously as

$$|\psi'\rangle = |\uparrow_1\rangle \otimes \left(|\uparrow_2\downarrow_3\rangle - |\downarrow_2\uparrow_3\rangle\right) \frac{1}{\sqrt{2}} \tag{14.8}$$

which is precisely analogous to (but the Hermitian conjugate of) the top half (the "bra") of the left hand side of Fig. 14.2.

It is easy to check that the inner product of these two states $|\psi\rangle$ and $|\psi'\rangle$, corresponding to the value of the left diagram of Fig.14.2 is[16]

$$\langle\psi'|\psi\rangle = -1/2$$

By redefining the normalization of these states, we can arrange for this overlap to have unit magnitude. However, the sign cannot be removed. The situation is the same for any two half-odd-integer spins fused to a singlet.

[16] This result of $-1/2$ is precisely the $6j$ symbol

$$\left\{ \begin{array}{ccc} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \end{array} \right\}$$

14.5 Appendix: The Isotopy Invariant Calculus

As mentioned in section 18.1.2 there is an alternate method of handling Frobenius-Schur indicators which allows one to generate knot invariants that are fully isotopy invariant — although it comes at the cost of not having a fixed gauge, which is unappealing for physical applications.

While one can apply this scheme to all the particles in the theory, for non-self-dual particles, and self-dual particles with postive Frobenius-Schur indicators, it is easy enough to just choose a gauge so that you can remove wiggles as in Fig. with $\epsilon = +1$. As such here we focus here only on self-dual particles with negative Frobenius-Schur indicators where the problem arises.

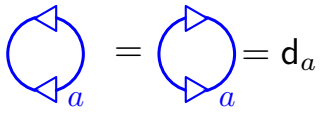
The scheme here is based on the idea that one should differentiate between a and $\bar{a}$ even if $a = \bar{a}$. In other words we should introducing

a new, artifical, degree of freedom which does not have any physical meaning at the end of the day. A version of this scheme is described by Kitaev [2006] which we will follow.

We assign each cup and cap a big triangular arrow (called a "flag") which can either point left or right. (These are not the same arrows we have been using to distinguish particles from their antiparticles. Indeed, the particles we are focusing on here are self-dual!). These flags are supposed to indicate which lines are an a particle going up versus a $\bar{a}$ particle going down (despite the fact that these two things are supposed to be the same!). The convention of Kitaev is that the flag always points towars $\bar{a}$.

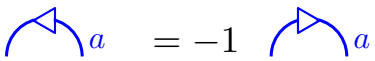
Hermitian conjugation does not flip the direction of the arrow

The corresponding inner product is positive definite so that



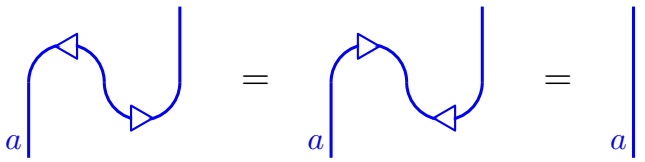
$$= \mathsf{d}_a$$

The direction of a flag can be reversed at the price of a factor of the Frobenius-Schur indicator (which is -1 since these are the only particles we are concerned with here)



$$= -1 \qquad \kappa_a = -1 \tag{14.9}$$

Finally, the cups and caps can cancel in pairs if the flags point in opposite directions, such as



(14.10)

Now, if we are given a diagram, without flags drawn this does not have a value (the factors of -1 are ambiguous). We must state how the flags are to be added to the the cups and caps of the diagram (for all self-dual particles with negtive Frobenius-Schur indicator) before the diagram gets a value.

If we make the rule that all flags always point right, this is basically fixing a unique gauge for all cups and caps. Making this choice, we recover the story of section 18.1.2 — that is removing a wiggle in a line incurs a minus sign (this can be seen from Eqs. 14.9 and 14.10).

However, another possible choice is to say that flags should be put on

cups and caps so that they alternate directions as you walk along any line. There are two ways that any line could be decorated with alternating flags, but fortunately, the two decorations are entirely equivalent[17]. Because wiggles with alternating flags can be freely straightened we then have an isotopy invariant diagrammatic set of rules.

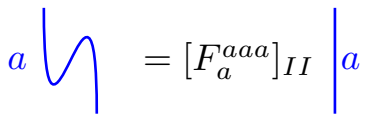
This scheme then provides a way that *any* braided anyon theory can be converted into an knot (or link) invariant with full isotopy invariance. Note, however, that it provides a different output from our above "gauge-fixed" choice. Also note (entirely analogous to note 15 above) if one tries to evaluate a link by using F-moves, one goes from a link diagram to a diagram with fusions — and these may not have full isotopy invariance. Further we comment that perhaps the easiest way to keep track of the flags in any scheme is to use isotopy first, but then before applying F-moves, flip flags using Eq. 14.9 to get get to a gauge fixed diagram (say with all flags pointing right) where signs are then easy to keep track of.

For many mathematics applications (certainly for knot theory) one wants to work with an isotopy invariant set of rules as described here. However, for many physics purposes, the gauge fixed set of rules will be more appropriate. The sign associated with a wiggle, *can* be a genuine physical quantity, as we described in section 14.4.

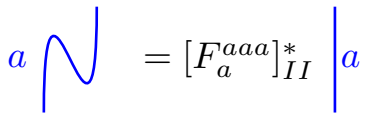
[17] The fact that the two possible decorations give the same result is a reflection of the fact that a and $\bar{a}$ are actually the same and so you can switch them and still get the same result. Recall that the flags are meant to point from a to $\bar{a}$ where these two are artificially declared to be different. However, switching which is which must not matter.

14.6 Appendix: $[F_a^{aaa}]_{II}$ is real

Let a be a self-dual particle (i.e., $a = \bar{a}$). Working with the physics normalization we already showed (Fig. 14.2) that

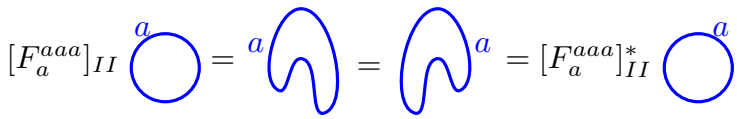


Similarly, using an inverse F-move, and the fact that F is unitary (See section 9.5.2) we derive



Equivalently the last diagram can be derived as being the Hermitian conjugate the previous diagram.

Finally, assuming only causal isotopy invariance, the equality



then shows that $[F_a^{aaa}]_{II}$ must be real.

14.7 Appendix: Some Additional Properties of Unitary Fusion Categories

Unitary fusion categories (the theories we have been discussing!) have two useful properties which we now present. bWe do not prove these properties here. More detailed discussion is given by Kitaev [2006]. More detailed discussions are given for example in Jones and Penneys [2017] or Etingof et al. [2005]. These latter references are quite mathematical.

14.7.1 Pivotal Property

A property that may seem obvious is known as the pivotal property. This states that there should be isomorphisms[18] between a vertex with a downturned line and one with an upturned line, such as that shown in Fig. 14.17. While this seems like a rather small statement (which is a

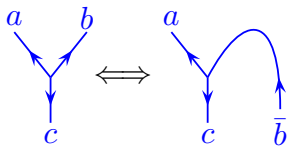


Fig. 14.17 A theory is pivotal if there exist isomorphisms between the states of the Hilbert spaces described by pairs of vertices that differ by downturning and upturning lines.

property of any unitary fusion category) it turns out to be quite powerful. One can deduce from this that the transformations in Figs. 14.10 and 14.11 are unitary — meaning that the constants on the right hand side have unit magnitude

$$\left|\sqrt{\frac{d_a d_c}{d_b}}[F_a^{c\bar{c}a}]_{Ib}\right| = \left|\sqrt{\frac{d_a d_c}{d_b}}[F_a^{a\bar{c}c}]_{bI}\right| = 1 \tag{14.11}$$

See Kitaev [2006]; Bonderson [2007]. In the more generael case where the fusion multiplicity N_{ab}^c is greater than one, the vertices have additional indices μ and ν and the transform is a unitary matrix in these indices. An example of this is given in Fig. 14.18.

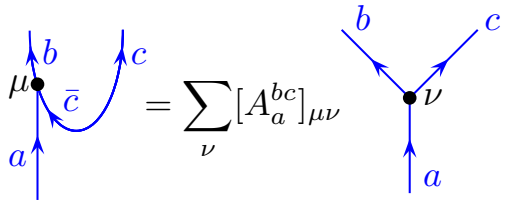


Fig. 14.18 The matrix A_a^{bc} is a unitary matrix in the indices μ and ν. In the simpler case of Fig. 14.11, the prefactor is a unitary one-by-one matrix, meaning it is a magnitude one complex scalar.

From this type of identity one successively turn up and down legs at

[18] We do not say there is a unitary tranformation between the two diagrams since the two diagrams operate on different Hilbert spaces —- the left diagram having one down leg and two up, whereas the right has one up and two down.

vertices to obtain the identity in Fig. 14.19 where the prefactor C in the figure is also a unit magnitude complex scalar (or a unitary matrix in the case where the vertex has an additional index).

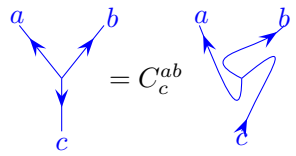


Fig. 14.19 The relationship between these two diagrams is unitary, meaning C is a just a phase. For cases where $N_{ab}^c > 1$ the vertices are marked with an index, say μ on the left and ν on the right, and for fixed a, b, and c, the constant C_c^{ab} becomes a unitary matrix $[C_c^{ab}]_{\mu\nu}$ in the indices μ, ν.

Quite a few more identities can also be derived from the pivotal property. Detailed discussions of this property (and its meaning) are given by Kitaev [2006]; Bartlett [2016]. One particularly useful identity is given by applying Fig. 14.19 three times in a row to obtain

$$C_c^{ab} C_a^{bc} C_b^{ca} = 1 \tag{14.12}$$

which diagrammatically is drawn as the so-called "pivotal identity" in Fig. 14.20. In the case where there are additional indices at the vertex, Eq. 14.12 becomes a matrix product which equals the identity matrix.

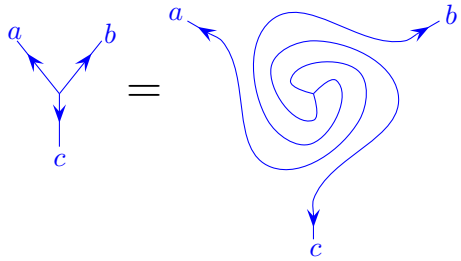


Fig. 14.20 The pivotal identity

The derivation of the pivotal identity is a bit complicated and is given by Kitaev [2006]. However, it can be made a bit more intuitive physically by turning up one of the branches to obtain the alternate form of the pivotal identity shown in Fig. 14.21. This form can be understood as the statement that the vacuum (or particles fusing to the vacuum) can be rotated freely in space-time.

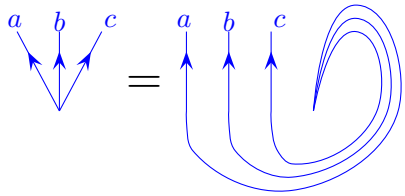


Fig. 14.21 Another version of the pivotal identity. We can derive this from Fig. 14.20 by turning up the c-leg.

14.7.2 Spherical Property

Theories which are unitary (describing real quantum mechanical particles) have an additional property called being "spherical". Given a diagram X with a line coming out the top and a line coming in the bottom. The so-called left trace is defined by connecting up the top line with the bottom line in a loop going to the left, as in the left of Fig. 14.22. The right trace is defined similarly, except that the loop goes to the right of the diagram X as in the right of Fig. 14.22. If the left trace is always equal to the right trace we say that the theory is spherical. The name here comes from the idea that we could pull the string around the back of a sphere in order to turn a left trace into a right trace as shown in Fig. 14.23. However, the spherical property is actually stronger than Fig. 14.23 suggests since it allows us to turn a right trace into a left trace even when there are other objects on the sphere which might prevent us from dragging a string all the way around the back of the sphere.

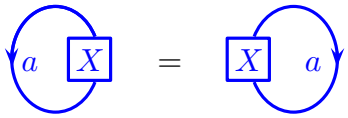


Fig. 14.22 The Spherical Property sets the left trace equal to the right trace as shown in the picture.

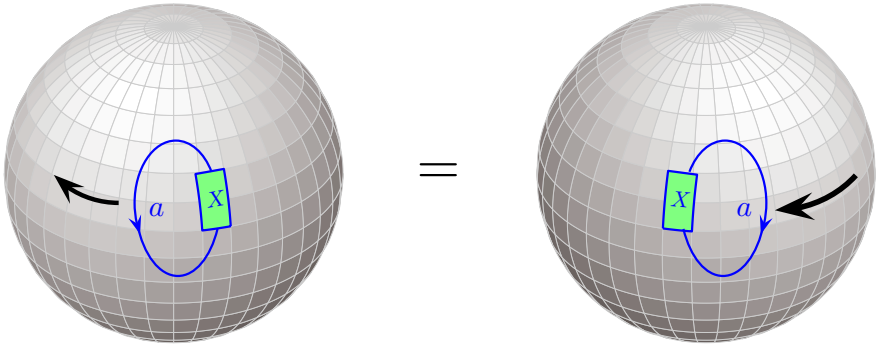


Fig. 14.23 The naming "spherical" comes from the idea that we can pull the string around the back of a sphere (as indicated by the black arrows) to turn a left trace into a right trace.

An obvious result of the spherical property is that $d_a = d_{\bar{a}}$.

14.8 Appendix: Higher Fusion Multiplicities

When we have a theory with higher fusion multiplicities (i.e., $N_{ab}^c > 1$ for at least one fusion channel), then the vertices must be given indices.

This appendix is identical to that of section 12.5 except that here we have changed the normalization from physics normalization to isotopy invariant normalization.

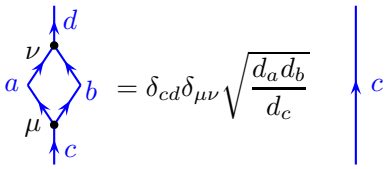


Fig. 14.24 The bubble diagram when there are fusion multiplicities. This diagram is drawn in the isotopy invariant normalization. Compare to Fig. 12.39.

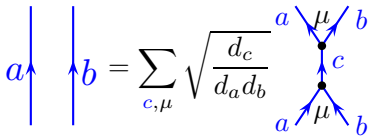


Fig. 14.25 Insertion of a complete set of states. When there are fusion multiplicities, these must be summed over as well $\mu \in N_{ab}^{c}$. This diagram is drawn in the isotopy invariant normalization. Compare to Fig. 12.40.

Further Reading

This is some reading.

Exercises

15

Twists

Recall from chapter 2 that we considered the procedure of pulling tight a loop in a ribbon as shown in Fig. 15.1 (compare Fig. 2.7). Pulling tight results in a twisted ribbon, which (viewing time as going vertically) corresponds to a particle twisting around its own axis, while at a fixed point in space, as time progresses. The twist in the ribbon can be removed at the cost of a complex phase which we call θ_a, known as the particle's **twist factor**. In other words, the particle twisting around its own axis accumulates a phase θ_a compared to a particle that does not twist. This phase is something we should expect for any particle with a spin, since rotating a spin accumulates a phase.

In our diagrammatic notation, we do not draw ribbons. Rather to represent a particle twisting around itself we use blackboard framing as discussed in section 15. We thus formally define the twist factor as given in Fig. 15.2.

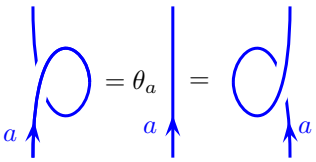


Fig. 15.2 The definition of the twist factor θ_a drawn using blackboard framed diagrams. The looped strings should be thought of as ribbons lying in the plane as in Fig. 15.1 which are equivalent to a ribbon that twists around its own axis.

Invoking the Hermitian conjugation principle (if we reflect a diagram around a horizontal axis, and reverse the arrows so they remain pointing in the same direction, we complex conjugate its amplitude, see Fig. 12.4) we similarly have the mirror image diagrams shown in Fig. 15.3

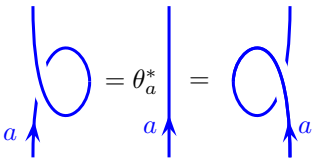


Fig. 15.3 The mirror image diagrams to those of Fig. 15.2.

It is easy to confirm that the twist factor θ_a can only be a unit magnitude phase[1] as expected (proof of this is given in Fig. 15.4).

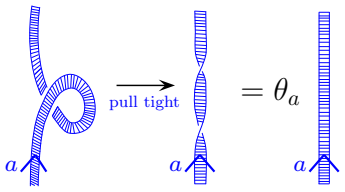


Fig. 15.1 Pulling tight a loop in a ribbon results in a twist. This twist in a ribbon of particle type a can be removed at the cost of a phase factor of θ_a. See also Fig. 2.7.

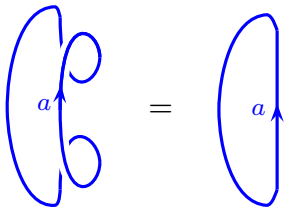


Fig. 15.4 This equality establishes $\theta_a\theta_a^* = 1$, hence $|\theta_a| = 1$. We can evaluate the diagram on the left by removing the two twists and getting $\theta_a\theta_a^*$ times the diagram on the right. On the other hand, the diagram on the left can also be turned directly into that on the right just by using moves which we know are allowed such as Fig. 13.8. (See also exercise 15.1).

[1] One might think that θ_a has been acting strange for a while — but it is just a phase. Ha ha!.

One further interesting fact about the twist factor is that

$$\theta_a = \theta_{\bar{a}}$$

which can be seen from the equality of the diagrams shown in Fig. 15.5.

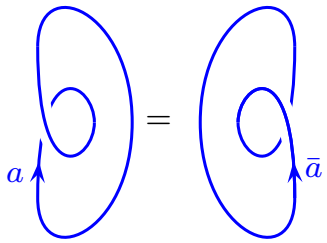


Fig. 15.5 The equaltiy of these diagrams establishes $\theta_a = \theta_{\bar{a}}$. See exercise 15.1.

The twist factor is related to the so-called **topological spin**, or **conformal scaling dimension**, usually called h_a, via the relation

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

This phase accumulated from a 2π rotation is what we typically get in quantum mechanics from the operator $e^{2\pi i \hat{S}}$ with $\hat{S}$ the spin operator and we set $\hbar = 1$. The vacuum, or identity particle, should have zero scaling dimension, $h_I = 0$.

Note that in many quantities of interest will depend only on the twist factor θ_a, i.e., the fractional part of the topological spin, h_a mod 1. Indeed, we will see that many of the topological properties of a system are independent of the integer part of the topological spin, and care only about the fractional part. That said, in chapter *** below we will also find cases where the integer part of h_a is important too.

Recall also the famous spin-statistics theorem (as discussed near Fig. 2.7), which tells us that the twist factor should give us the phase for exchanging two identical particle, and is thus intimately related to the anyonic statistics of particles. Of course two cases are very well known to us: if the spin h_a is an integer, then $e^{2\pi i h_a}$ is the identity, and the particle is a boson. If h_a is a half-odd-integer, then the phase is -1 and the particle is a fermion.

15.1 Relations between θ and R

Braiding and twisting are very closely related to each other. In fact, twist factors θ are related to the R-matrices we introduced in chapters 10 and 13 in several different ways.

First, let us try to evaluate the looped ribbon in Fig. 15.2 using the R-matrix as in Fig. 15.6. This manipulation establishes the relation[2,3]

$$\theta_a = \sum_c \frac{\mathsf{d}_c}{\mathsf{d}_a} R_c^{aa} \tag{15.1}$$

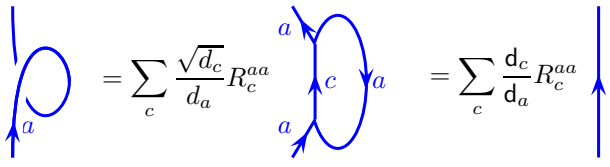


Fig. 15.6 Relation of the twist factor to the R-matrix. In the first step we use Fig. 14.9. In the second step we use Fig. 14.11 along with Eq. 14.11 and finally Eq. 14.7. For $d_a < 0$, we have implemented rule 0 from section 14.2.1 in the last step.

[2] In cases where $N_{aa}^c > 1$, then R_c^{aa} is replaced by $\sum_\mu [R_c^{aa}]_{\mu\mu}$.

[3] Note that θ_a is gauge independent, whereas R_c^{ab} is generally gauge dependent. (Although R_c^{aa} is gauge independent as well).

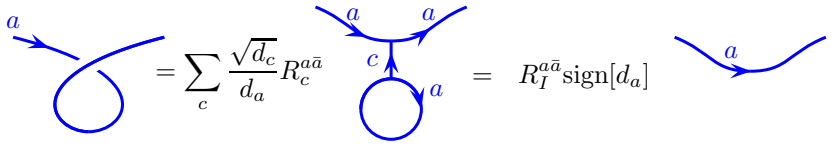


Fig. 15.7 The first step invokes Fig. 14.9. In the second step we realize that c must be the identity I by locality with $d_c = d_I = 1$. The remaining a-loop cancels a factor of d_a against $|d_a|$ (the loop evaluates to a positive number for any unitary theory) leaving only the sign. Note that this diagram is not precisely equal to θ_a^* (See Fig. 15.8).

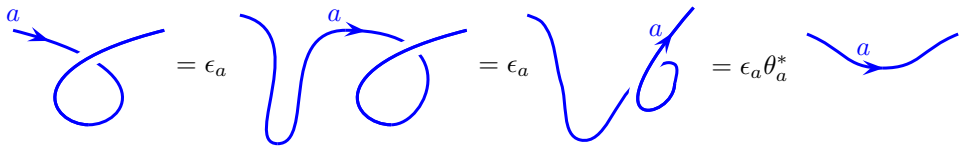


Fig. 15.8 A loop in rope turned sideways gets a twist factor θ^* along with a wiggle factor ϵ. The first step introduces a wiggle in the curve and we incur a factor of ϵ_a as in Fig. 14.5. The second step is an allowed smooth deformation (See exercise 15.1). In the last step we remove the loop and obtain a factor of θ_a^* as in Fig. 15.3.

A second, and different, relationship can be derived via the manipulations shown in Fig. 15.7. One might be tempted to identify the left of Fig. 15.7 with the twist factor θ but this is not quite right when we look at it carefully, as shown in Fig. 15.8. From Fig. 15.7 and 15.8 we derive[4]

$$R_I^{a\bar{a}} = \epsilon_a \mathrm{sign}[d_a]\, \theta_a^* = \mathrm{sign}([F_a^{a\bar{a}a}]_{II})\, \theta_a^* \quad . \tag{15.2}$$

where we have used Eq. 14.1 to relate ϵ, d_a, and F. Recall in particular that for self-dual particles $[F_a^{a\bar{a}a}]_{II}$ is the Frobenius-Schur indicator.

The final relationship between R and θ is known as the ribbon identity[5]

$$R_c^{ba} R_c^{ab} = \frac{\theta_c}{\theta_a \theta_b} \tag{15.3}$$

which can be derived by the geometric manipulations in Fig. 15.9.

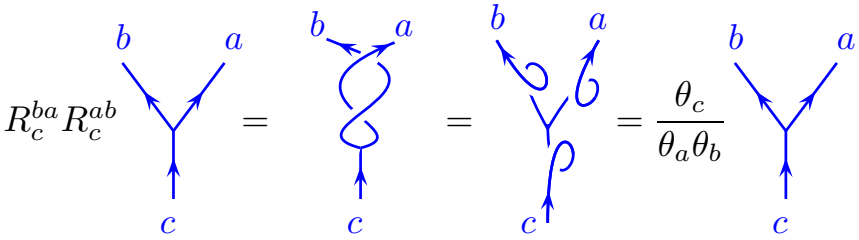


Fig. 15.9 Deriving the ribbon identity. The middle is the nonobvious geometric step. See also exercise 15.2.

[4] Note that in this equation, and Figs. 15.6 and 15.7 we are assuming the gauge choice discussed in section 14.2.1.

[5] This is also gauge independent.

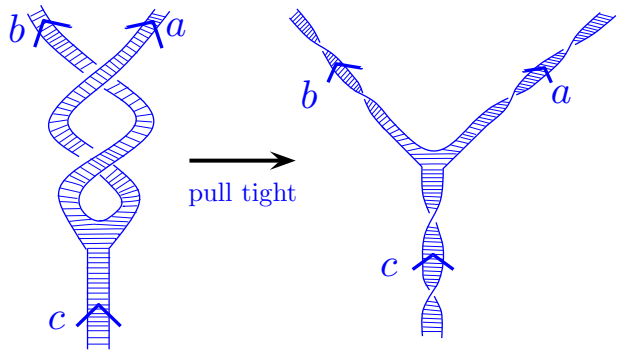


Fig. 15.10 The middle step of Fig. 15.9 viewed as a ribbon diagram.

The middle step in Fig. 15.9 is perhaps non-obvious, but is perhaps clarified if viewed as a ribbon diagram as in Fig. 15.10. See also exercise 15.2.

Exercises

Exercise 15.1 Using Geometric Moves I
(a) Using the allowed moves in Fig. 13.8, show the equivalence of the left and right of Fig. 15.4 (b) Similarly, show the equivalence of the left and right of Fig. 15.5. (c) Similarly show the equivalence of the middle two figures in Fig. 15.8.

Exercise 15.2 Using Geometric Moves II
Demonstrate the middle step of Fig. 15.9 by using allowed geometric moves such as Fig. 13.8 and Fig. 13.9 and Fig. 13.13. You may also need the pivotal identity Fig. 14.20.

Exercise 15.3 Gauge Independence of Ribbon Identity
Show that the ribbon identity Eq. 15.3 is gauge independent.

Theories with Full Isotopy

16

In chapters 12 through 15 we carefully developed the principles of anyon diagrammatics. In the current chapter we aim for a slightly simplified and abbreviated, but still extremely useful, version of the diagrammatic rules developed (roughly axiomatically) in the prior chapters.

Our original intent for a TQFT was to develop rules that would map a labeled knot or link diagram into a complex amplitude output (as in Fig. 7.1) in a way that would be invariant under any smooth deformations (isotopy) of space-time. Most generally in topological theories, we found that there could some restrictions on what sort of deformations of space-time would leave the output unchanged (See for example section 14.1.1). In the current section we will focus on a simpler class of theories where these impediments are lifted. In particular the topological theories of this chapter have the property that they give the same output amplitude for *any* smooth deformation of space-time (treating world lines as ribbons). In other words, in this chapter we assume our theories have "full isotopy" invariance (or "regular isotopy" invariance as discussed in section 2.6.1). An example of such full isotopy is shown in Fig. 16.1.

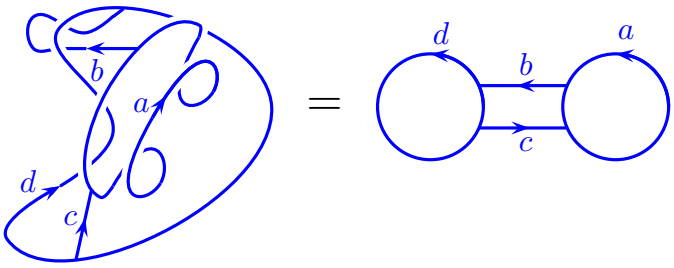


Fig. 16.1 For a theory with full isotopy invariance (regular isotopy invariance) these two diagrams must evaluate to the same result since one can be continuously deformed into the other treating the lines as ribbons.

16.1 Planar Diagrams

We start by considering only planar diagrams, so we do not allow over- and under-crossings (which we re-introduce in section ***). Because of our specialization to these fully isotopy invariant theories, our rules for diagrammatic manipulation will be slightly easier than those in chapter 12.

As in chapter 12 there is still a bra and ket interpretation of diagrams. Roughly we can think of cutting a diagram in half and viewing one side as a bra and the other as a ket[1]. We can also roughly think of these diagrams as being world lines of particles moving in 1+1 dimensions.

[1]We can think of any direction as being time, although it is sometimes most convenient to think of time as up.

16.1.1 Planar Diagrammatic Rules

In this chapter all (regular) isotopy of lines is allowed. In particular we are freely allowed to make the deformation shown in Fig. 16.2. In the language of Fig.14.5 we are assuming all $\epsilon_a = +1$.

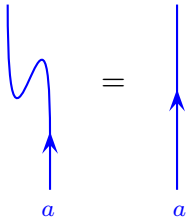


Fig. 16.2 For the isotopy invariant theories considered in this chapter, this deformation is allowed.

As in previous chapters we would like to use F-matrices to help us convert one diagram into another. Although we previously found that bending lines up and down (as in Fig. 14.10) can incur nontrivial factors, in this chapter we instead assume no such nontrivial factors so we may turn up and down legs freely. Our F-matrix can thus be written as in Fig. 16.3. Note that the conventions we use in this chapter are different from that of the previous chapter but instead match those introduced by Levin and Wen [2005].

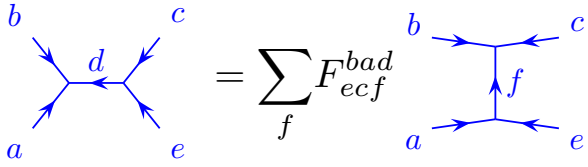


Fig. 16.3 The definition of the F-matrix for fully isotopy invariant theories. This notation uses the conventions of Levin and Wen [2005]. For a unitary theory the F-matrix with fixed indices a, b, c, d is unitary in the indices d and f. For this F matrix to be nonzero, the vertices in the pictures must be allowed fusions — i.e., $N_{abd} = N_{ced\bar{}} = N_{bcf} = N_{ae\bar{f}} = 1$. The case with fusion multiplicities N greater than one is considered in section 16.4.

In this chapter, the orientiation of this diagram (how we direct the legs compared to some direction we call time) does not matter. Further, we can freely rotate the diagrams in Fig. 16.3 and we can bend legs up and down freely as well. For example, the same F-matrix as in Fig. 16.3 applies to Fig. 16.4.

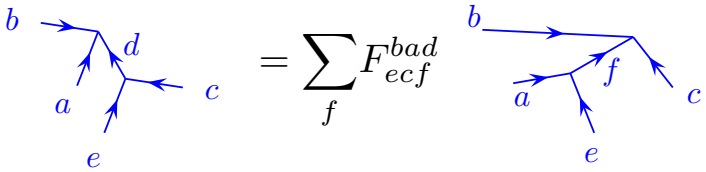


Fig. 16.4 For fully isotopy invariant theories, the F-moves can be deformed in arbitrary ways. For example the same F matrix governs the transformation in Fig. 16.3 as in this figure.

We can compare the definition of F-matrix in Fig. 16.3 to our prior definition of the F-matrix shown in Fig. 9.1. Since we now assume that

we can bend legs up and down freely, we can bend legs in Fig. 9.1 and reverse arrows to make it look like Fig. 16.3 and we thereby derive the relation between the two definitions

$$F_{ecf}^{bad} = [F_e^{\bar{a}\bar{b}\bar{c}}]_{df} \tag{16.1}$$

Again the idea of the F-matrix is to write a single diagram (on the left of Eq. 16.3) as a sum of diagrams on the right. By successively applying such F-moves to parts of complicated diagrams we can restructure any given diagram in a multitude of ways.

There are several further useful rules for diagram evaluation. First, we need is to give a value to the a labelled loop as in Fig. 16.5. As in the case of the Kauffman bracket invariant, the value of the loop will be called d, although here there will be a different d_a, called the "quantum dimension", for each possible particle type a. Note that we have not yet shown the relationship between this definition of the quantum dimension d_a and the definition of d_a (which we also called "quantum dimension") given in Eq. 3.8. In section 17.1 we will show that these two definitions are in fact the same up to a possible sign!

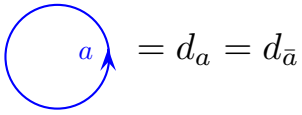


Fig. 16.5 The value of a loop labeled a is given by the quantum dimension d_a. Here we have invoked the spherical assumption to give us $d_a = d_{\bar{a}}$.

It is always true that $d_I = 1$, meaning that loops of vacuum can be freely added or removed from a diagram. As emphasized in section 14.1, giving the loop this normalization implies we are working with non-normalized kets (see Fig. 14.3, and also note 18 of chapter 2). Note that the theories we consider in this chapter must satisfy

$$\text{sign}[d_a]\text{sign}[d_b] = \text{sign}[d_c] \quad \text{whenever} \quad N_{ab}^c > 0 \tag{16.2}$$

as we described previously in section 14.2.1.

Secondly we define the contraction of a bubble as shown in Fig. 16.6.

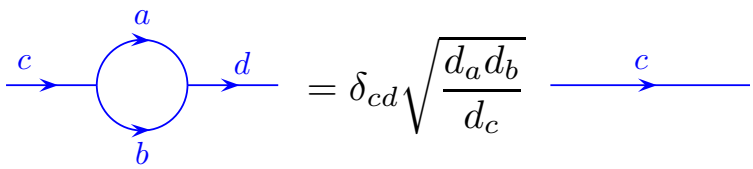


Fig. 16.6 Contraction of a bubble for fully isotopy invariant theories. In cases where some d's are negative we interpret the sign outside the square root as negative if and only if both $d_a < 0$ and $d_b < 0$.

This identity is the same as Fig. 14.7 only written sideways (in this

chapter the orientation of the diagram on the page does not matter). Physically we should think of this as being a version of the locality rule of section 8.2 — looked at from far away, one does not see the bubble. In particular this locality rule implies the "no-tadpole" rule, that any diagram of the sort shown in Fig. 16.7 must vanish unless the incoming line is the vacuum.

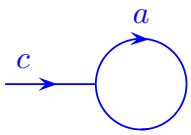


Fig. 16.7 Picture of a tadpole. (Apparently this picture is supposed to look like a tadpole.). The locality principle Fig. 16.6 implies that any diagram containing a tadpole must vanish unless the incoming line is labeled with the vacuum. (I.e., unless there is no incoming line!). Famously, Physical Review did not allow the use of the name "spermion" for diagrams of this sort.

We also have again the completeness relation as shown in Fig. 16.8.

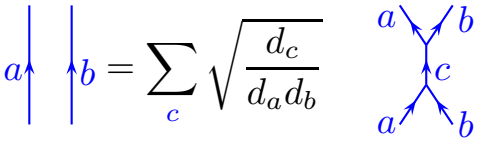


Fig. 16.8 Insertion of a complete set of states. In cases where some d's are negative we interpret the sign outside the square root as negative if and only if both $d_a < 0$ and $d_b < 0$.

This relation is precisely the same as Fig. 14.8, only now we can orient the diagram in any direction.

16.1.2 Summary of Planar Diagram Rules For Fully Isotopy Invariant Theories

Given the rules established in section 16.1.1, we can evaluate any planar diagram[2] and turn it into a complex scalar number made up of factors of F's and d's— very similar to what we did with the Kauffman bracket invariant, only without over- and under-crossings here. Here are a summary of the rules for diagram evaluation in the case of fully isotopy invariant planar theories. These rules are analogous to those presented in section 12.4, only here the rules are simpler.

[2]Any planar diagram with no loose ends. As described in detail in section 12.1 a diagram with loose ends should be considered a bra or ket or operator.

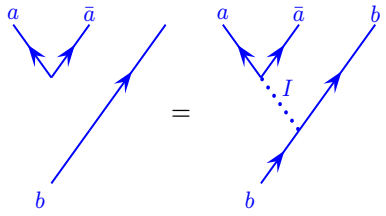


Fig. 16.9 One can always add or remove the identity (or vacuum) line to any diagram.

(1) One is free to continuously deform a diagram in any way as long as we do not cut any strand (for this section we assume no over- or under-crossings).

(2) One is free to add or remove lines from a diagram if they are labeled with the identity or vacuum (I). See the example in Fig. 16.9.

(3) Reversing the arrow on a line turns a particle into its antiparticle (See Fig. 8.4).

(4) A line must maintain its quantum number unless it fuses with another line, or splits.

(5) Vertices are allowed for multiplicities $N_{ab}^{c} > 0$ (See section 8.3). This includes particle-antiparticle creation and annihilation processes where $N_{a\bar{a}}^{I} = 1$ (an example is shown in Fig. 16.9).

(6) One can use F-moves to change the structure of diagrams.

(7) One can use relations Fig. 16.8 and 16.6 to change the structure of diagrams.

(8) Every diagram can be reduced to a set of loops which can each be evaluated to give d_a for each loop of type a.

16.1.3 Constraints and Examples

There are many constraints on our diagrammatic algebras for planar isotopy invariant theories. Here we give such constraints and explain where they all come from.

Constraint: The Pentagon

The consistency condition on F-matrices given in Eq. 9.7 can be converted to the notation of this chapter (See Eq. 16.1) to give[3]

$$F_{edl}^{c\bar{f}g} F_{e\bar{l}k}^{baf} = \sum_{h} F_{gch}^{baf} F_{edk}^{\bar{h}ag} F_{kdl}^{cbh} \tag{16.3}$$

[3] In deriving Eq. 16.3 from Eq. 9.7 we have taken $a, b, c, d \to \bar{a}, \bar{b}, \bar{c}, \bar{d}$ for ease of notation.

Constraint: Relating F to d

For any theory with full planar isotopy the value of d_a should be fixed by the F-matrices:

$$d_a = \frac{1}{F_{\bar{a}aI}^{\bar{a}aI}} \tag{16.4}$$

This is demonstrated by the manipulations of Fig. 14.4, converted into the notation of the current chapter. Recall that we are assuming $\epsilon_a = +1$.

Constraint: Inversion

One can perform an F-move on the right hand side of Fig. 16.3 to bring it back into the form on the left. We obtain the diagrammatic relation shown in Fig. 16.10,

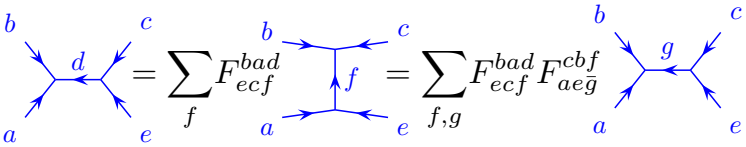


Fig. 16.10 In the second step we apply the same F-matrix equation from Fig. 16.3, but the diagram is rotated by 90 degrees.

which necessarily implies the consistency condition

$$\sum_f F^{bad}_{ecf} F^{cbf}_{ae\bar{g}} = \delta_{dg} \tag{16.5}$$

Constraint: Rotation

Rotating the diagram in Fig. 16.3 by 180 degrees and comparing it to the original diagram, one derives

$$F^{bad}_{ecf} = F^{ec\bar{d}}_{ba\bar{f}} \quad . \tag{16.6}$$

Constraint: Turning Up and Down

For a theory to be fully isotopy invariant, we must be able to freely make the moves shown in Fig. 14.10. As shown there, this requires $1 = \sqrt{(d_a d_c)/d_b}[F_a^{c\bar{c}a}]_{Ib}$, or in the notation of this chapter

$$F^{c\bar{c}I}_{a\bar{a}b} = \sqrt{\frac{d_b}{d_a d_c}} \tag{16.7}$$

whenever $b \times c = a + \ldots$, with the sign of the square root taken negative if and only if d_a and d_c are both negative.

Constraint: Unitarity

As mentioned above in Fig. 16.3, the F-matrix, being a change of basis, must be unitary. This means that

$$\sum_f F^{bad}_{ecf} [F^{bad'}_{ecf}]^* = \delta_{dd'} \tag{16.8}$$

$$\sum_d F^{bad}_{ecf} [F^{bad}_{ecf'}]^* = \delta_{ff'} \tag{16.9}$$

or equivalently $[F^{ba}_{ec}]^\dagger = [F^{ba}_{ec}]^{-1}$ Comparing the former to Eq. 16.5 we obtain

$$[F^{bad}_{ecf}]^* = F^{cbf}_{ae\bar{d}} \tag{16.10}$$

Constraint: Hermitian Conjugation

Using reflection across the horizontal axis as in Fig. 12.4, we can reflect the F-matrix equation Fig. 16.3 and compare the reflected to the unreflected diagram to obtain

$$F^{bad}_{ecf} = [F^{\bar{a}\bar{b}\bar{d}}_{\bar{c}\bar{e}f}]^* \tag{16.11}$$

$$= [F^{\bar{c}\bar{e}d}_{\bar{a}\bar{b}f}]^* \tag{16.12}$$

$$= F^{\bar{e}\bar{a}f}_{\bar{b}\bar{c}d} \quad , \tag{16.13}$$

where in the second line the first line has been used in combination with Eq. 16.6, whereas in the third line the first line has been used in

combination with Eq. 16.10.

Constraint: Reflection

An independent condition that is very often imposed is that the F-matrix should be invariant under left-right reflection. Compare the diagram shown in Fig. 16.11 to that of Fig. 16.3.

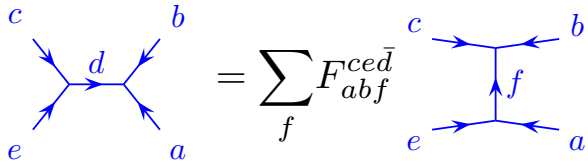


Fig. 16.11 The diagrammatic equation in Fig. 16.3 after being left-right reflected.This is necessary for a isotopically invariant 2+1 dimensional theory, but is an independent assumption for a planar diagram algebra.

If a theory has left-right reflection symmetry, then we must have a further constraint

$$F^{bad}_{ecf} = F^{ce\bar{d}}_{abf} \tag{16.14}$$

While this additional condition is not required for a planar diagram algebra, and one can even have full isotopy invariance in two dimensions without it, it is often assumed. For isotopically invariant three dimensional theories, such a symmetry is necessary since one can view the diagrams either from the front or the back.

Using Eq. 16.14 along with Eq. 16.13 gives us the natural seeming constraint

$$F^{bad}_{ecf} = [F^{\bar{b}\bar{a}\bar{d}}_{\bar{e}\bar{c}\bar{f}}]^* \tag{16.15}$$

Example: Evaluating a bubble

As an example of showing how further constraints are derived, let us use F-moves to evaluate the bubble shown in Fig. 16.12.

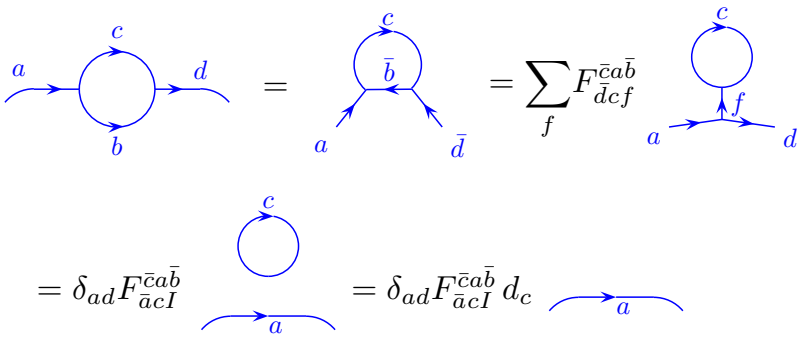


Fig. 16.12 Evaluation of a bubble diagram. In the first step, as usual we can flip the direction of an arrow and turn a particle into its antiparticle. In the second step we apply an F-move (compare to Fig. 16.3). Then by the no-tadpole (locality) rule (Fig. 16.7) , we can set f to the vacuum particle I and hence $a = d$.

However, we also know the value of the diagram in Fig. 16.12 from Fig. 16.6 which gives us $\sqrt{d_c d_b/d_a}$. Thus we derive $F^{\bar{c}a\bar{b}}_{\bar{a}cI} d_c = \sqrt{\frac{d_c d_b}{d_a}}$, or equivalently (while replacing b with $\bar{b}$ for simplicity and using $d_b = d_{\bar{b}}$) we have

$$F^{\bar{c}ab}_{\bar{a}cI} = \sqrt{\frac{d_b}{d_a d_c}} \tag{16.16}$$

whenever $c \times b = a + \ldots$ where the sign of the square root is taken negative if and only if d_a and d_b are both negative. Note that Eq. 16.16 could also be obtained from Eq. 16.7 with Eq. 16.13.

Example: The Theta diagram

A commonly considered diagram is the Theta diagram $\Theta(a,b,c)$ shown in Fig. 16.13. This diagram is easily evaluated by using Fig. 16.6 along with the value of a single bubble Fig. 16.5.

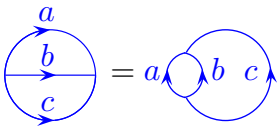
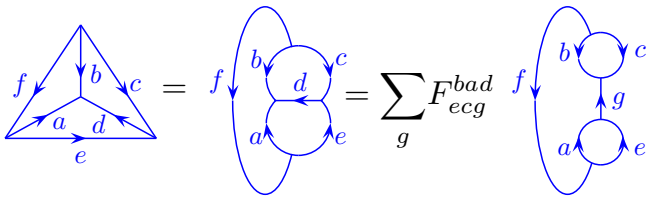
$\Theta(a,b,c) =$  $= d_c \sqrt{\frac{d_a d_b}{d_c}} = \sqrt{d_a d_b d_c}$

Fig. 16.13 The Theta diagram. This is evaluated by using Fig. 16.6 along with the value of a single bubble Fig. 16.5. The sign on the square root is taken negative unless all three d_a, d_b and d_c are positive.

Example: The tetrahedral diagram

Let us consider one more evaluation known as the tetrahedral diagram as shown in Fig. 16.14. At this point we are considering this as a planar diagram even though it looks three dimensional! However, we usually consider diagrams to be well defined if they live on the surface of a sphere, so if we want to think about this as being three dimensional, we should think of this as living on a spherical surface.



$$= F^{bad}_{ecf}\, d_f \sqrt{\frac{d_b d_c}{d_f}} \sqrt{\frac{d_a d_e}{d_f}} \equiv G^{bad}_{ecf}$$

Fig. 16.14 Evaluation of the tetrahedral diagram. The first step is just smooth deformation. The second step is application of an F move. Using Fig. 16.6, the index g must be equal to the index f and we obtain some factors of $\sqrt{d}$. Finally we are left with a single loop of f which gives a factor of d_f to give the final result which we give the name G. As in Fig. 16.6 the square roots are taken negative if and only if both d's in the numerator of the square root are negative.

For theories with full planar isotopy, the tetrahedral diagram has some obvious symmetries. For example, we should have rotational symmetry in the plane as shown in Fig. 16.15 which implies the identity (note the

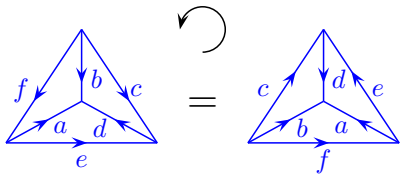


Fig. 16.15 An obvious rotational symmetry of the tetrahedral diagram.

definition of G in Fig. 16.14)

$$G^{bad}_{ecf} = G^{dba}_{f\bar{e}\bar{c}} \quad . \tag{16.17}$$

Another symmetry comes from Eq. 16.6

$$G^{bad}_{ecf} = G^{ec\bar{d}}_{ba\bar{f}} \quad . \tag{16.18}$$

which we draw as shown in Fig. 16.16.

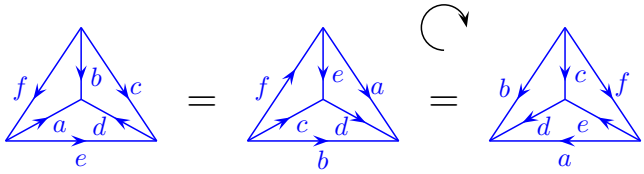


Fig. 16.16 The first step is the identity in Eq. 16.18 and the second step is a rotation as in Fig. 16.15. Although this is actually a planar diagram it appears as a rotation in 3D.

Although the diagram shown in Fig. 16.16 is a planar diagram, from the far left to the far right, it appears as if it is a rotation in 3D. Using Fig. 16.15 and 16.16 we can rotate this tetrahedron in any way we like. If one assumes the reflection symmetry Eq. 16.14 then one can also take the mirror image of the tetrahedron as well to obtain an equivalence between 24 tetrahedral diagrams related by symmetries[4].

[4] For an example of a spherical category that cannot be put in a form with full tetrahedral symmetry, see Hong [2009].

16.2 Braiding Diagrams Revisited

So far in this chapter we have considered planar theories only. Extension to fully (regular) isotopy invariant three dimensional[5] theories follows almost exactly the expositions of chapters 13. Here we will recapitulate the key points.

[5] Here we mean 2+1 dimensional theories, but we sometimes may not specify a particular time direction.

First, any regular isotopy (See section 2.6.1) of diagrams is allowed and does not change the value of the diagram. This means that as long as we treat the lines in a diagram as ribbons, we can deform the diagram into any shape we like. Often this sort of regular isotopy can turn a diagram with braiding into a planar diagram which can then be evaluated using the rules of section 16.1.2. An example of this is shown in Fig. 16.1.

Most generally, however, we will not be able to eliminate all over- and under-crossing of lines just by using isotopy (i.e., by deforming a diagram). To handle crossings, we invoke the R-matrix discussed in chapter 13. The basic moves we need are summarized in Fig. 16.17 (which just repeat results previously discussed in Fig. 14.9). In the current chapter, where we consider fully isotopy invariant theories, the orientation of the crossing does not matter. So, for example, the same R-matrix formula applies to the crossing in Fig. 16.17 (top) as in Fig. 16.18.

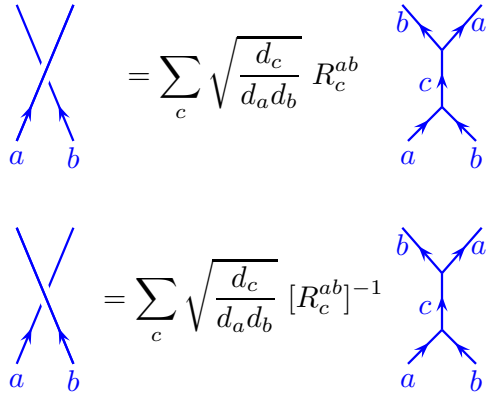


Fig. 16.17 Resolving a crossing with isotopy normalization. The square roots are taken negative if any only if d_a and d_b are both negative

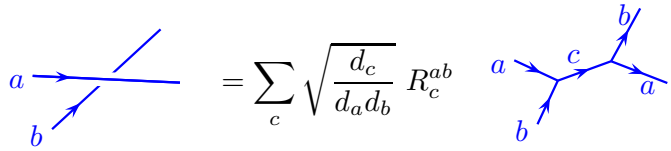


Fig. 16.18 With fully isotopy invariant theories, we can rotate diagrams freely. Thus the uncrossing formula here is identical to that in Fig. 16.17 (top).

The R-matrix moves in Fig. 16.17 allow us to take any diagram with over-and under-crossings and turn it into a planar diagram[6] which can then be evaluated using the rules of section 16.1.2.

We thus add two rules for evaluation of diagrams in three dimensions to our previously stated rules for planar diagrams of section 16.1.2:

(1') One is free to continuously deform diagrams in three dimensions in any way as long as we do not cut any strands (and strands are treated as ribbons). In other words we have *regular isotopy* of diagrams (See section 2.6.1). [This rule replaces rule (1) from the list in section 16.1.2]

(9) Over- and under-crossings can be turned into planar diagrams using the R-matrix as in Fig. 16.17

[6] Turning a crossing into a planar diagram is known as "resolution of a crossing".

16.2.1 Constraints

There are several further constraints on the braiding diagrammatic algebra which we now mention.

Constraint: Rotation

As mentioned in Fig. 16.18, we can rotate crossings freely. By turning a crossing entirely upside-down and comparing to the original crossing, we obtain an identity which holds for isotopy invariant theories

$$R_c^{ab} = R_{\bar{c}}^{\bar{b}\bar{a}}. \tag{16.19}$$

Constraint: Hermitian Conjugation

As in Eq. 13.1 we can use Hermitian conjugation to derive

$$[R_c^{ab}]^{-1} = [R_c^{ba}]^* \tag{16.20}$$

Constraint: Hexagon Equations

As discussed in section 13.3 there are consistency conditions between F and R matrices known as hexagon equations (Eqs. 13.2 and 13.3). In the notation of the current chapter these can be written as

$$R_e^{ca}\, F_{d\bar{b}g}^{\bar{c}\bar{a}e}\, R_g^{cb} = \sum_f F_{d\bar{b}f}^{\bar{a}\bar{c}e}\, R_d^{cf}\, F_{d\bar{c}g}^{\bar{b}\bar{a}f} \tag{16.21}$$

$$[R_e^{ca}]^{-1} F_{d\bar{b}g}^{\bar{c}\bar{a}e}\, [R_g^{cb}]^{-1} = \sum_f F_{d\bar{b}f}^{\bar{a}\bar{c}e}\, [R_d^{cf}]^{-1}\, F_{d\bar{c}g}^{\bar{b}\bar{a}f} \tag{16.22}$$

Relation to Twists

As detailed in chapter 15 each particle has a twist factor $\theta_a = \theta_{\bar{a}}$ (with $\theta_I = 1$) describing twisted strands as chapter 15 which we show again here for completeness in Fig. 16.19.

Fig. 16.19 Definition of Twist Factors (See chapter 15 for more details)

In the current chapter the direction the twist is drawn on the page is not important (only its chirality)

The R-matrix and the twist factors are related in several ways. The identities Eq. 15.1, 15.2, and 15.3 hold where for fully isotopy invariant theories we set $\epsilon_a = +1$ for all particles.

16.3 Negative d_a and Unitarity

We have allowed theories with d_a to be negative. While this may not seem problematic, if we think of a loop as being an inner product $\langle \text{state}|\text{state}\rangle$ as in Fig. 14.3, a negative d_a implies a non-positive-definite inner product, which is forbidden in quantum mechanics. This apparent problem is discussed in detail in sections 14.1–18.1.2. Here we will very briefly adapt the scheme discussed in section 14.2.1 to the current situation.

In short, we accept that our diagrammatic algebra has negative d_a's. Evaluation of such diagrams with negative d_a's we call the *non-unitary* evaluation. However, with a small reinterpretation of the diagrams we can still think of these diagrams as describing a unitary theory. As discussed in section 14.2.1 we add two simple rules to our list.

(0') We must break the space-time symmetry and define a time direction (often up on the page).

(0) Before evaluating a diagram, count the number of negative-d caps, and call it n. After fully evaluating the diagram multiply the final result by $(-1)^n$.

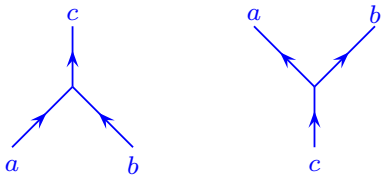


Fig. 16.20 With time going vertical, the left diagram is a neagtive-d cap if and only if $d_a < 0$ and $d_b < 0$. (The directions of the arrows do not matter, and if the particles are self-dual we do not draw arrows). The right diagram is never a negative-d cap.

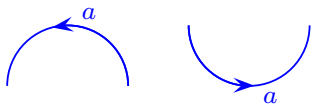


Fig. 16.21 With time going vertical, the left diagram is a negative-d cap if and only if $d_a < 0$. The right diagram is never a negative-d-cap. We can think of these diagrams as being the same as the diagrams in Fig.16.20 with c being the identity. The directions of the arrows do not matter.

Recall from section 14.2.1. Here a negative-d cap occurs when we go forward in time and two particles with $d < 0$ come together to annihilate or form a particle having $d > 0$. (See examples in Figs. 16.20 and 16.21).

These modifications guarantee we are describing a unitary theory. For example, if we take as simple loop like Fig. 16.5 with $d_a < 0$ the naive evaluation (before application of rule 0) gives a negative result. However, the diagram has one negative-d cap and so the result is multiplied by $(-1)^1$ thus giving a postive result as we should expect for a diagram that can be interpreted as $\langle \text{state}|\text{state}\rangle$ as in Fig. 14.3. More examples of how these evaluations work are given in section 14.2.1.

The situation described in this section — having a theory which is fully isotopy invariant but has negative d_a — is quite common. Fundamentally, as discussed in sections 14.1–18.1.2, the need to use negative d comes from negative Frobenius-Schur indicators. There are many topological theories of this type — including very simple theories the semion theory $SU(2)_1$ and more generally theories like $SU(2)_k$.

16.4 Appendix: Higher Fusion Multiplicities

As in section 9.5.3 when fusion multiplicities are greater than one, the vertices have additional indices which we label with greek indices $\mu, \nu, \ldots$. For example, if a and b fuse to c with $N_{ab}^c > 1$ then the vertex will have an additional index $\mu \in 1 \ldots N_{ab}^c$. Note that compared to section 9.5.3 we do not put black dots on the vertices here. In the conventions of the current chapter we would then have

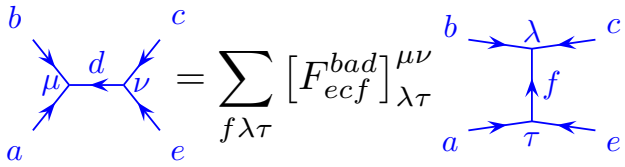


Fig. 16.22 *F*-matrix for isotopy invariant theories with fusion multiplicity.

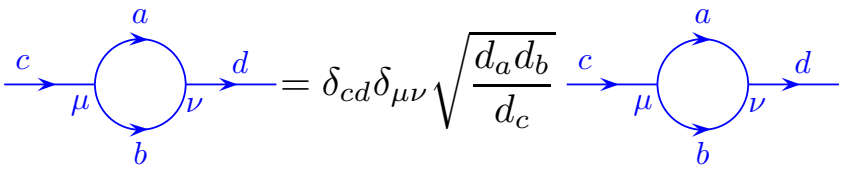


Fig. 16.23 The locality principle for the isotopy invariant diagrammatic algebra with fusion multiplicity.

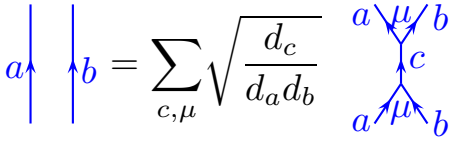


Fig. 16.24 Insertion of a complete set of states with isotopy normalization and fusion multiplicity.

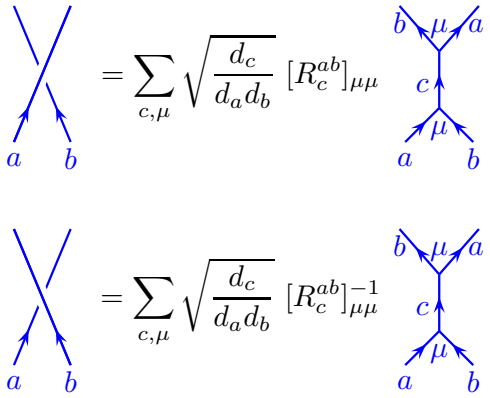


Fig. 16.25 Resolving a crossing with isotopy normalization and fusion multiplicity. The square roots are taken negative if any only if d_a and d_b are both negative

Chapter Summary

- THis is an item

Further Reading

This is some reading.

Levin-Lin?

Summary of properties of category?

17 Further Structure

In this chapter we will explore some further structure that is inherent in topological theories.

17.1 Quantum Dimension

Recall that we defined d_a, the quantum dimension, in terms of how fast the Hilbert space grows as we fuse together many a particles (See Eq. 3.8)

$$\text{Dim of } M \text{ anyons of type } a \sim \mathsf{d}_a^M$$

An alternative definition (Eq. 8.9) is that[1]

$$\mathsf{d}_a = \text{the largest eigenvalue of the matrix } N_a \tag{17.1}$$

[1] Recall that N_a is defined as the matrix with components $[N_a]_b^c = N_{ab}^c$ which are the fusion multiplicities.

We have claimed several times that these quantum dimensions are (up to a possible sign) equal to the value of a loop d_a in our diagrammatic algebra. In this section we will finally prove this important result.

To make a connection to our diagrammatic algebra, let us consider fusing two loops labeled a and b as shown in Fig. 17.1[2]

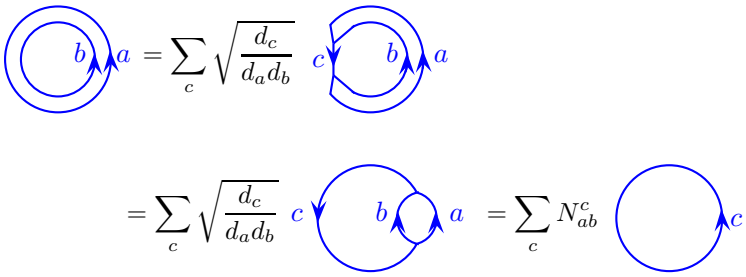


Fig. 17.1 Fusing two loops into a single loop. In the first line we use the completeness relation Fig. 14.8, then we deform to the second line and finally in the last step we remove the bubble using Fig. 14.7.

The result seems rather natural, that a and b can fuse together to form

[2] This result holds very generally. There are several possible worries one might have about this calculation which we should dispell. First, in cases where $N_{ab}^c > 1$ one must conisder additional indices at the vertices, in which case we use Fig. 14.25 and Fig. 14.24 in place of Fig. 14.8 and Fig. 14.7 in the derivation. Secondly, one might worry that for general theories, without full isotopy invariance, going from the first line to the second line might be problematic. However, it turns out that one does not need full isotopy invariance, just the pivotal property is enough to get to the second line (See section 14.7.1 and exercise 17.1).

c in all possible ways. The derivation uses the completeness relation in the first line (Fig. 14.8), then we deform to get to the second line, and finally in the last step we remove the bubble using Fig. 14.7.

Now the value of a loop in our diagrammatic algebra is d_a. However, if d_a is negative we must implement rule 0 from section 14.2.1 and we always obtain a positive definite result $|d_a|$ for a single loop (which is appropriate for a diagram which can be interpreted as $\langle$state|state$\rangle$, see Fig. 14.3). Thus the diagramatic manipulations of Fig. 17.1 give us[3]

$$|d_a|\ |d_b| = \sum_c N_{ab}^c\ |d_c| \tag{17.2}$$

[3]One says that the quantities $|d_a|$ form a representation of the fusion algebra. Compare to Eq. 8.3.

Eq. 17.2 has an interesting interpretation if we define a vector $\vec{d}_{abs}$ to have components $|d|_c$ with the vector index being c. We can then rewrite the Eq. 17.2 as an eigenvalue equation

$$|d|_a\ \vec{d}_{abs} = N_a\ \vec{d}_{abs}$$

Since N_a is a matrix of nonnegative numbers[4], and $\vec{d}_{abs}$ is a vector of positive numbers, we identify $\vec{d}_{abs}$ as the so-called Perron-Frobenius eigenvector of the matrix N_a, and its eigenvalue $|d|_a$ is guaranteed by the Perron-Frobenius theorem (see appendix 17.7) to be the largest[5] eigenvalue of N_a which, by Eq. 17.1 gives us

$$\mathsf{d}_a = |d|_a$$

as claimed. This derivation does not rely on the theory having a well-defined braiding.

[4]The Perron-Frobenius theorem is usually applied to matrices of all positive numbers and is slightly weaker when applied to the nonnegative case. See the detailed discussion in appendix 17.7.

[5]Strictly speaking N_a may have several eigenvalues of the same absolute magnitude with the Perron-Frobenius eigenvalue being the only one which is real and positive. This does not change our conclusion.

17.2 The unlinking $\tilde{S}$-matrix

First let use the locality principle (or no-transmutation) principle (See Fig. 8.7) to show that a closed loop of type a around a world line of type x gives some constant which we call $\tilde{S}_{ax}$ as shown in Fig. 17.2. (In fact using the R-matrix we can explicitly derive this identity and evaluate $\tilde{S}_{ax}$ in terms of twist factors θ, the fusion multiplicities N_{ax}^c and the quantum dimensions d. See exercise 17.3. However, we will not need this explicit expression.) Note in particular that $\tilde{S}_{Ix} = 1$ since the identity loop can be removed for free, and $\tilde{S}_{aI} = \mathsf{d}_a$ since a single loop[6] of a gives d_a.

[6]Here if d is negative we include the sign associated with rule 0 of section 14.2.1 in the evaluation of the diagram so that we obtain a postive quantum dimension d.

Now, 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.17.3. This identity is entirely analogous to that of Fig. 17.1. In essence we are fusing a and b to form all possible strands c and then in the last step we apply $\tilde{S}$.
On the other hand, we could also evaluate the left hand side of Fig. 17.3 by applying the identity of Fig. 17.2 twice in a row, as in Fig. 17.4.

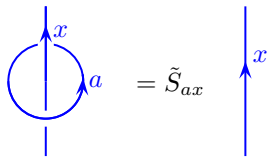


Fig. 17.2 The locality principle tells us that the value of a loop of a around a world line x is some number which we call $\tilde{S}_{ax}$. (Indeed, we can use the R-matrix to calculate $\tilde{S}_{ax}$. See exercise 17.3.)

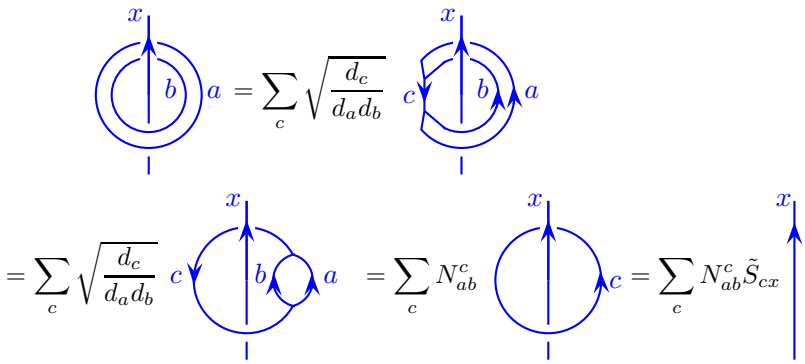


Fig. 17.3 Similar reasoning as in Fig. 17.1 allows us to write this diagrammatic relationship.

Equating the result of Fig. 17.3 to that of Fig. 17.4 we obtain

$$\tilde{S}_{ax}\tilde{S}_{bx} = \sum_c N_{ab}^c \tilde{S}_{cx} \tag{17.3}$$

This result holds for any anyon theory with a well defined braiding (i.e., that satisfies the hexagon relationship). Note that in the special case where x is the identity we just recover Eq. 17.2.

17.3 The (modular) S-matrix

Recall from section 7.3.1 that we defined the S-matrix (Eq. 7.6) in several ways. On the one hand we defined

$$S_{ab} = Z(S^3; a \text{ loop linking } b \text{ loop}) \tag{17.4}$$

whereas on the other hand, we said that S was (under certain conditions that the theory has no transparent particles) a unitary transformation between two different bases for describing the Hilbert space of a torus.

Let us more generally define the S-matrix in the following way

$$S_{ab} = \quad a \quad b \quad \times \frac{1}{\mathcal{D}} \tag{17.5}$$

where $\mathcal{D}$ is the so-called *total quantum dimension*[7]

[7] Since our rules for evaluating a diagram (Eq. 14.6) tell us to interpret a diagram as $Z(S^3; \text{link diagram})/Z(S^3)$, comparing Eq. 17.4 to Eq. 17.5 we can conclude that $Z(S^3) = 1/\mathcal{D}$.

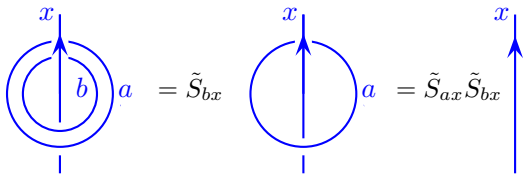


Fig. 17.4 Application of $\tilde{S}$ twice (compare to Fig. 17.2).

$$\mathcal{D} = +\sqrt{\sum_a |\mathsf{d}_a|^2} \tag{17.6}$$

In chapter *** we will show that this correctly corresponds to a basis transform matrix as we described in section 7.3.1 above.

Using isotopy of diagrams and Hermitian conjugation it is easy to establish (see exercise 17.2) that

$$S_{ab} = S_{ba} = S_{\bar{a}\bar{b}} = S_{\bar{b}\bar{a}} = S^*_{\bar{a}b} = S^*_{b\bar{a}} = S^*_{a\bar{b}} = S^*_{\bar{b}a} \tag{17.7}$$

And further by setting one of the indices to the vacuum I, we are left with a single loop that evaluates to the quantum dimension[8], hence giving us

$$S_{Ia} = S_{aI} = \mathsf{d}_a/\mathcal{D}. \tag{17.8}$$

[8]Here again we are using rule 0 from section 14.2.1 so that a single loop always evaluates to a positive number $\mathsf{d}_a = |d_a|$ independent of the sign of d_a.

The normalization constant $\mathcal{D}$ here is chosen so that

$$\sum_a |S_{aI}|^2 = 1$$

which is required if the S-matrix is to be unitary. Note in particular that this implies

$$S_{II} = 1/\mathcal{D} \tag{17.9}$$

Let us now evaluate the S-matrix in terms of our unlinking matrix $\tilde{S}$. By bending the top of x in Fig. 17.2 and forming a closed loop with the bottom of x, we construct linked rings as shown in Fig. 17.5.

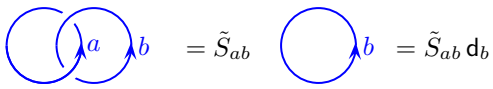


Fig. 17.5 Evaluation of linked rings. In the case where $d_b < 0$, we have applied rule 0 from section 14.2.1 so that the single loop gives us a positive d_b quantum dimension.

Comparing to the definition of S in Fig. 17.5 we obtain

$$S_{ab} = \tilde{S}_{ab}\,\mathsf{d}_b/\mathcal{D} = \tilde{S}_{ab}S_{Ib}$$

or equivalently

$$\tilde{S}_{ab} = \frac{S_{ab}}{S_{Ib}}$$

Plugging this into Eq. 17.3 gives us

$$\frac{S_{ax}S_{bx}}{S_{Ix}} = \sum_c N_{ab}^c S_{cx}. \tag{17.10}$$

Again this is generally true for any braided anyon theory, i.e., any theory which satisfies the pentagon and hexagon relations.

17.3.1 Unitary S = Modular

When S is unitary, we say the theory is *modular*[9]. It turns out that S is unitary if and only if the identity is the only *transparent* particles in the theory[10]. A particle a is said to be transparent if braiding a all the way around any other particle accumulates no phase. Equivalently, in terms of the S-matrix we can write

$$a \text{ is transparent} \quad \Leftrightarrow \quad S_{ax} = \frac{\mathsf{d}_a \mathsf{d}_x}{\mathcal{D}} \qquad \text{for all } x \tag{17.11}$$

In other words, a full braiding of a with any particle is trivial[11].

It is clear that an S matrix cannot be unitary if there is any transparent particle a besides the identity (since the rowd S_{ax} and the row S_{Ix} would be proportional to each other). What is not obvious is that the absence of any transparent particle guarantees S is unitary. This statement can be proven using our axiomatic diagrammatic principles (i.e., not invoking any of the topological discussion of section 7.3.1 above). However, the proof is a bit complicated and we refer the reader to Kitaev [2006]; Etingof et al. [2015] for details.

As mentioned a number of times, in some sense all "well-behaved" anyon theories are modular (we say they are *modular tensor categories*). Unfortunately, there are common theories which are not modular. For example, a simple theory of a single fermion which obtains a minus sign under exchange. A full braiding gives a plus sign exactly like the vacuum, and is hence non-modular (although it may have a well defined braiding). It is usually the case that theories containing fermions are non-modular.

Let us assume for the remainder of this section that we have a modular theory, i.e., that S is unitary, as we had stated in section 7.3.1 above will discuss this assumption more in a moment), we can multiply Eq. 17.10 by $S^{-1} = S^\dagger$ on the right to obtain the often quoted Verlinde formula[12]

$$N_{ab}^c = \sum_x \frac{S_{ax}S_{bx}[S^{-1}]_{xc}}{S_{Ix}} = \sum_x \frac{S_{ax}S_{bx}S_{cx}^*}{S_{Ix}} \tag{17.12}$$

which tells us that all the information about the fusion algebra is contained entirely within the S matrix!.

Alternatively, one can multiply Eq. 17.10 by $S^{-1} = S^\dagger$ on the left to

[9] We will explain the meaning of the word "modular" in section 17.3.2.

[10] We ran into this concept as far back as section 4.3.2

[11] An equivalent statement using the ribbon identity Eq. 15.3 is that a is transparent if and only if $\theta_c/(\theta_a\theta_b) = 1$ for all b and c whenever $N_{ab}^c \neq 0$.

[12] Verlinde [1988] derived this in the context of conformal field theories. In different context it was derived earlier by Pasquier [1987].

obtain (with S and N treated as matrices on the left)

$$[S^\dagger N_a S]_{xy} = \delta_{xy} \left(\frac{S_{ax}}{S_{Ix}} \right)$$

This means that the S matrix (at least for modular theories) is the unitary matrix that simultaneously diagonalizes all of the N_a fusion multiplicity matrices (We mentioned this previously after Eq. 8.12).

A useful quick application of the Verlinde formula, Eq. 17.12, is to write an expression for the conjugation[13] matrix

$$C_{ab} \equiv \delta_{a\bar{b}}$$

This is simply a permutation matrix that permutes each particle with its antiparticle. Obviously $C^2 = \mathbf{1}$ is the identity.

We can find a relationship between C and the S-matrix by writing C as the fusion multiplicty matrix of two particles fusing to the identity

$$C_{ab} = N^I_{ab} = \sum_x S_{ax} S_{bx} = [SS^T]_{ab}$$

where we have used the Verlinde formula to evaluate N^I_{ab} along with S_{Ix} being real. Finally using the fact S is symmetric we obtain

$$C = S^2$$

[13]The word "conjugation" is meant to evoke charge conjugatoin which changes positive charges to negative charges.

17.3.2 The Modular Group and Torus Diffeomorphisms

Let us define one more matrix, which is the diagonal matrix of the twist factors

$$\tilde{T}_{ab} = \theta_a \delta_{ab}$$

It turns out to be more useful to absorb an additional complex phase into this matrix, so let us define

$$T = \tilde{T} e^{-2\pi i c/24} \tag{17.13}$$

where c is a real constant, known as the *chiral central charge* which is an important piece of data for an anyon theory[14]. The central charge modulo 8 can be calculated from the twist factors and quantum dimensions via (Fröhlich and Gabbiani [1990]; Rehren [1990]),

$$e^{2\pi i c/8} = \frac{1}{\mathcal{D}} \sum_a \mathsf{d}_a^2 \, \theta_a \tag{17.14}$$

[14]We will run into the central charge c again in section *** where we make connection between TQFTs and 1+1 dimensional conformal field theories. In a 1+1 dimensional chiral conformal field theory, the specific heat is given in terms of the central charge as $c_v = \pi k_B^2 T c/(3v)$ where v is the speed of light, T is temperature, k_B is Boltzmann's constant π is pi and 3 is three.

The set of operations generated by T and S form[15] a group known as the *modular group.*[16]

$$S^2 = C \qquad C^2 = \mathbf{1} \qquad (ST)^3 = C \tag{17.15}$$

[15]We do not need C as an independent generator since $C = S^2$.

[16]Be warned there are several closely related groups that are sometimes known as the modular group.

For proof of the last identity we again refer the reader to Kitaev [2006]; Etingof et al. [2015]. These relations are equivalent to the group $SL(2,\mathbb{Z})$, the group of two-by-two matrices with integer coefficients and unit determinant which has generators

$$\underline{S} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \qquad \underline{T} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

It is easy to check that Eq. 17.15 are satisfied by these two by two matrices.

The modular group has a beautiful topological interpretation: it is the group of topologically distinct[17] orientation preserving diffeomorphisms of the torus surface. To see how this works we consider a torus to be a plane $\mathbb{R}^2$ with the lattice of integers $\mathbb{Z}^2$ modded out so that each lattice point is identified with every other lattice point.

$$T^2 = \mathbb{R}^2/\mathbb{Z}^2$$

Any transformation on the plane that one-to-one maps lattice-points to lattice points gives a representative diffeomorphism of the torus. Such transformations are given by the elements of the group $SL(2,\mathbb{Z})$ just by mapping points $\vec{v}$ in the plane to $A\vec{v}$ where A is an member of $SL(2,\mathbb{Z})$[18].

This analogy with the diffeomorphisms of the torus is certainly not coincidental! Let think back to the discussion of section 7.3 and 7.4. We considered the solid torus $D^2 \times S^1$ with a particle world line of type a around the handle. We wrote the wavefunction on the surface as

$$|\psi_a\rangle = |Z(D^2 \times S^1; a)\rangle$$

which forms an orthonormal basis $\langle\psi_a|\psi_b\rangle = \delta_{ab}$. As mentioned there, this inner product corresponds to sewing together the two solid tori to create $S^2 \times S^1$. However, we could also have sewn the tori together after exchanging meridian and longitude to create S^3, and the inner product then becomes (See Eq. 7.6)

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

A different way of thinking of this is that we make a diffeomorphism on the surface of the torus before gluing the two halves back togther. Thus we could have said

$$\langle\psi_b|\hat{S}|\psi_a\rangle = S_{ab}$$

where $\hat{S}$ is the operator that makes the diffeomorphism on the surface (the diffeomorphism precisely exchanges meridian and longitude).

Note further that $\hat{C} \equiv \hat{S}^2$ exchanges the meridian and longitude twice, giving a net effect of rotating the torus by 180 degrees. If we think back to the solid torus with a world line of type a around the handle, the rotation of the torus surface by 180 degrees changes the relative direction of the embedded world line and thus changes a to $\bar{a}$, implementing the conjugation operation C.

[17] What we mean here is the so-called "mapping-class group" of the torus surface. I.e., two mappings of the torus surface to the torus surface are considered to be the same if one can be smoothly deformed into the other.

[18] It is obvious that such a transformation of the plane corresponds to a diffeomorphism of the torus. What is a bit less obvious is that *all* diffeomorphisms of the torus are topologically equivalent to (i.e, can be smoothly deformed into) a linear map of this sort. See, for example, Rolfson [1976]; Farb and Margalit [2012] for detailed discussions of this point.

Finally, the $\hat{T}$ operation on the torus surface implements an elementary Dehn twist as described in section 7.4. The fact that the T operation on the state $|\psi_a\rangle$ corresponds to a twist factor θ_a is fairly obvious from looking at Fig. 7.15. The presence of the additional complex factor related to the central charge in Eq. 17.13 is a subtle point and stems from a change in the 2-framing discussed in section 5.3.4.

17.4 Periodic Tables of TQFTs

As mentioned in sections 9.3 and 13.3, anyon theories (or TQFTs) are extremely constrainted by the pentagon and hexagon equations. Indeed, for any given set of fusion rules there are only a finite number of possible solutions up to gauge equivalence (Etingof et al. [2005]). Once one includes additional conditions, such as the theory being modular, the number of possible solutions drops even more. This makes it possible to consider building a "periodic table" of possible TQFTs — i.e., a complete list of all consistent modular solutions of pentagon and hexagon. The procedure for building this table is to hypothesize that there are only n different particle types, with n a small number. With fixed n one can constrain the possible fusion rules, then pentagon solutions, and finally hexagon solutions. For modular tensor categories this program has been carried out by Rowell et al. [2009] for up to four particle types[19] Some of the key results from this periodic table is presented in table 17.1. In chapters *** we will see a few of the basic principles used for compiling such a table.

[19] See also earlier work by Gepner and Kapustin [1995] as well as Bonderson [2007].

Extensions of the idea of a periodic table have been made in a number of directions. Of note, the periodic table has been extended to include all possible braided theories (solutions of pentagon and hexagon) without imposing the modularity condition. Further since theories with fermions are such an important special case, theories that only fail to be modular due to a fermion, have been put in periodic tables as well. References to these works are given in the further reading below.

17.5 Ω Strand (Kirby Color)

A particularly useful object to consider is a weighted sum of particle types particular sum of particle types known as an Ω-strand, or sometimes "Kirby color" strand[20], as shown (purple) in Fig. 17.6.

[20] Sometimes the Ω strand is normalized with an additional factor of $\mathcal{D}^{-1}$ out front.

Fig. 17.6 A String of Kirby color (Ω-strand) is a weighted superposition of all anyon string types[20]. Note that the Kirby color string does not have an arrow on it since it is an equal sum over all particles and their antiparticles. Here d is the quantum dimension, S is the modular S-matrix, and I is the identity, or vacuum, particle.

This weighted sum will occur in several different contexts, including

Table 17.1 Rank≤ 4 modular tensor categories. From Rowell et al. [2009].

particle types	Fusion rules	# solutions	Examples
I		1	$(E_8)_1$
I, X	$X^2 = I$	2	$\mathbb{Z}_2 = SU(2)_1$ = Semion
I, X	$X^2 = I + X$	2	$(G_2)_1$=Fibonacci
I, X, Y	$X^2 = Y$, $XY = I$, $Y^2 = X$	2	$\mathbb{Z}_3 = SU(3)_1$
I, X, Y	$X^2 = I + Y$, $XY = X$, $Y^2 = I$	8	$SU(2)_2$, Ising, $(E_8)_2$
I, X, Y	$X^2 = I + X + Y$, $XY = X + Y$, $Y^2 = I + X$	2	$SU(2)_5/\mathbb{Z}_2$
I, X, Y, Z	$X^2 = Y = Z^2$, $XZ = I = Y^2$, $XY = Z$, $ZY = X$	4	$\mathbb{Z}_4 = SU(4)_1$
I, X, Y, Z	$X^2 = I$, $XY = Z$, $XZ = Y$, $Y^2 = I$, $YZ = X$, $Z^2 = I$	1 1 3	of type $D(\mathbb{Z}_2)$=Toric Code of type $(D_4)_1$ of type $\mathbb{Z}_2 \times \mathbb{Z}_2$
I, X, Y, Z	$X^2 = I + X$, $XY = Z$, $XZ = Y + Z$, $Y^2 = I$, $YZ = X$, $Z^2 = I + X$	4	Fibonacci × $\mathbb{Z}_2$
I, X, Y, Z	$X^2 = I + X$, $XY = Z$, $XZ = Y + Z$, $Y^2 = I + Y$, $YZ = X + Z$, $Z^2 = I + X + Y + Z$	3	Fibonacci × Fibonacci
I, X, Y, Z	$X^2 = I + X + Y$, $XY = X + Y + Z$, $XZ = Y + Z$, $Y^2 = I + X + Y + Z$, $YZ = X + Y$, $Z^2 = I + X$	2	$(G_2)_2$, $SU(2)_7/\mathbb{Z}_2$

Some comments on this table: G_2, E_8, D_4 and $SU(N)$ are Lie groups. Theories listed, such as $(G_2)_2$ are Chern-Simon theories with the last subscript indicting the level of the theory. $SU(2)_5/\mathbb{Z}_2$ means start with the Chern-Simons theory $SU(2)_5$ and consider only the "even" subset of anyons, which form a complete theory by themselves. (Similar for $SU(2)_7/\mathbb{Z}_2$). $D(\mathbb{Z}_2)$ is the quantum double of $\mathbb{Z}_2$ which we will discuss in section *** . The counting of the number of solutions in most cases is obvious. For $SU(2)_1, (G_2)_1, SU(2)_5/\mathbb{Z}_2$ and $(G_2)_2$ we have the right (R) and left (L) handed versions of the theory (The left and right version of $(D_4)_1$ turn out to be the same). The 3 cases of Fib × Fib and $\mathbb{Z}_2 \times \mathbb{Z}_2$ correspond to taking $L \times L$, $R \times R$, and $L \times R$. Note that $R \times L$ is equivalent to $L \times R$ so does not count as a different theory. However, for Fib $\times\mathbb{Z}_2$, all four $L \times L$, $R \times R$, $R \times L$ and $L \times R$ are different. We will discuss the 8 versions of anyon theories related to the Ising or $SU(2)_2$ in section *** below. Note that the $(E_8)_1$ theories need not be entirely trivial – even though it has no nontrivial particles – since it can have a central charge which is any integer multiple of 8. Note that nowhere in this table do we have fusion multiplicity $N_{ab}^c > 1$.

sections 22.3,***,*** below. The Ω-strand has some interesting properties. For example, in Fig. 17.7 we show the so-called "killing property" of a loop of Ω — a loop of Ω allows only the identity to go through it: all other particle types are "killed."

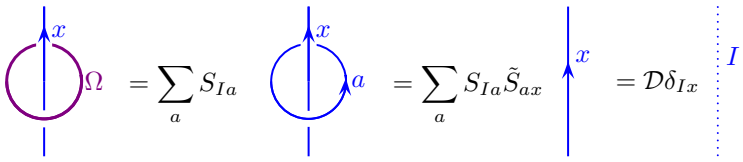


Fig. 17.7 The killing property of the Ω-strand for modular theories. A loop of Kirby color (Ω) allows no particles through it except the identity. In the final step we use unitarity of S (and the fact that $S_{Ia} = S^*_{Ia}$).

A useful corrolary of the Killing property is given in Fig. 17.8.

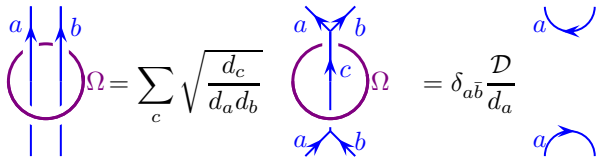


Fig. 17.8 The Ω strand joins two line due to the killing property which we use in the second step to force c to be the identity.

A loop of Ω strand with a twist, shown in Fig. 17.9, interestingly reproduces our expression for the central charge modulo 8, as in Eq. 17.14.

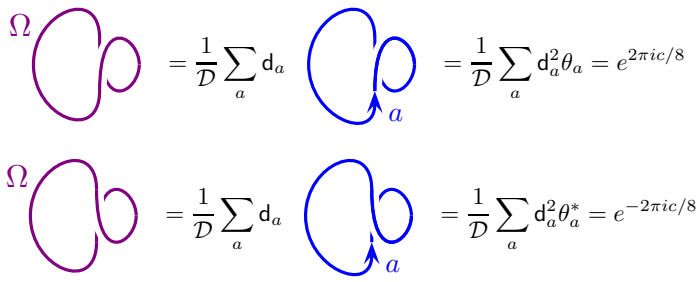


Fig. 17.9 For modular theories, a twisted Ω-strand gives $e^{\pm 2\pi i c/8}$ with c the (real) central charge. See Eq. 17.14. (See also exercise 22.5)

17.6 Still Further Structure

Modular (and many non-modular) theories have a great deal of extra structure that we have not even touched on. The theories are obviously very highly constrained, so it is rather natural to expect that there will be

many nontrivial relationships between the quantities we have discussed. A useful relationship which is assigned as exercise 17.3 is

$$S_{ab} = \frac{1}{\mathcal{D}} \sum_c N_{a\bar{b}}^c \frac{\theta_c}{\theta_a \theta_b} \mathsf{d}_c \quad . \tag{17.16}$$

Another interesting result is a theorem by Bantay [1997] which gives us the following nontrivial relationship between the Frobenius-Schur indicator κ_k of a particle k and the modular S matrix

$$\sum_{i,j} N_{ij}^k S_{0i} S_{0j} \frac{\theta_i^2}{\theta_j^2} = \begin{cases} \kappa_k = \epsilon_k \text{sign}(d_k) & k = \bar{k} \\ 0 & k \neq \bar{k} \end{cases} \tag{17.17}$$

A beautiful theorem by Vafa [1988] tells us that for any braided unitary theory (modular or not) all the spin factors θ_a must be an n^{th} root of unity so that $\theta_a^n = 1$ where the integer n is determined only by the fusion multiplicity matrices N_{ab}^c. For example, we have

$$\prod_b \theta_b^{X_{ab}} = 1 \tag{17.18}$$

with

$$X_{ab} = -2N_{a\bar{a}}^b N_{a\bar{a}}^{\bar{b}} - N_{aa}^b N_{\bar{a}\bar{a}}^{\bar{b}} + 4\delta_{ab} \sum_q N_{aa}^q N_{\bar{a}\bar{a}}^{\bar{q}}$$

17.7 Appendix: Perron-Frobenius Theorem

The Perron-Frobenius theorem states that for a matrix with all positive entries, there is a unique eigenvector with all positive entries (up to multiplication by an overall constant) known as the Perron-Frobenius eigenvector. The corresponding eigenvalue is the largest magnitude eigenvalue, and is positive. Further for a unitarily diagonalizable nonzero matrix with all non-negative entries, if there exists an eigenvector with all positive entries (up to multiplication by an overall constant), its eigenvalue is positive and is of magnitude greater or equal to all other eigenvalues. The application of the Perron-Frobenius theorem to the fusion matrix N_a is a bit tricky since the theorem is stronger when the matrix being considered has strictly postive entries and the N_a matrices are only guaranteed to have nonnegative entries. To avoid this problem, construct an arbitrary sum of the N_a matrices $M = \sum_a \alpha_a N_a$ where all the coefficients α_a are positive. Since all the $N_a's$ have common eigenvectors (since they are normal matrices and they commute with each other, see section 8.3.1), these are also the eigenvectors of the matrix M. Further, all the elements of M are strictly positive, so we may apply the Perron-Frobenius theorem for postive definite matrices to M. We thus obtain a Perron-Frobenius eigenvector of M with strictly positive entries (up to a multiplicative constant). But the eigenvectors of N_a match those of of M so we have a postive definite eigenvector of N_a which then must be the Perron-Frobenius eigenvectors whose eigenvalue

is greater or equal to any eigevalue of N_a.

17.8 Appendix: Algebraic Derivation of the Verlinde Form

In this section we show that we do not need the structure of braiding in order to derive an equation of the Verlinde form, analogous to Eq. 17.12. Let us begin by recalling from section 8.3.1 that for any topological theory[21,22] the fusion rules are described by fusion multiplicity matrices N_{ab}^c which can be viewed as a set of square normal matrices N_a with indices b and c. As discussed in section 8.3.1 these normal matrices commute with each other and therefore can be simultaneously diagonalized by a matrix which we will call U (See Eq. 8.12)

$$N_a = U\lambda^{(a)}U^\dagger \tag{17.19}$$

where $\lambda^{(a)}$ is a diagonal matrix for each a. We note again that we will discover below that for a so-called modular braided theory we will find that S is the modular S-matrix. More generally we call U the *mock* S-matrix.

From Eq. 17.19, the columns of U are the simultaneous eigenvectors of the N matrices which we can make explicit as

$$\sum_c [N_a]_b^c U_{cd} = U_{bd}\lambda_d^{(a)} \tag{17.20}$$

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 called the vacuum I (or identity) which fuses trivially with all other particles, we have $[N_a]_I^c = \delta_a^c$ so we have

$$U_{ad} = \sum_c [N_a]_I^c U_{cd} = U_{Id}\lambda_d^{(a)}$$

so that

$$\lambda_d^{(a)} = \frac{U_{ad}}{U_{Id}} \tag{17.21}$$

substituting back into Eq. 17.19 we get the Verlinde formula

$$[N_a]_b^c = \sum_x U_{bx}\frac{U_{ax}}{U_{Ix}}U_{cx}^* \quad . \tag{17.22}$$

This result is extremely general[22].

[21] Here we can mean any planar algebra (unitary fusion category) or any 2+1 dimensional topological theory with a braiding (unitary braided fusion category).

[22] All we actually need is a commutative fusion ring with a unique identity and inverses.

17.9 Appendix: Algebraic Derivation that Quantum Dimensions Form a Representation of the Fusion Algebra

Recall that the columns of U are the simultaneous eigenvectors of the N_a matrices. Invoking the Perron-Frobenius theorem, there must be a particular index z such that[23] the eigenvector U_{bz} has all postive entries with fixed z (up to an overall multiplicative constant). This is a common eigenvector of all the N_a matrices, and the corresponding eigenvalues are $\lambda_z^{(a)}$. By the Perron-Frobenius theorem, since the eigenvector is all positive $\lambda_z^{(a)}$ must be the largest eigenvalue of N_a. Recalling (Eq. 8.9) that the quantum dimension can also be defined as the largest eigenvalue of N_a, we have

$$\mathsf{d}_a = \lambda_z^{(a)}$$

Now let us multiply the Verlinde relation Eq. 17.22 on both sides by $\mathsf{d}_c = \lambda_z^{(c)} = U_{cz}/U_{Iz}$ (see Eq. 17.21) and sum over c. We have

$$\begin{aligned}\sum_c N_{ab}^c \mathsf{d}_c &= \sum_{x,c} U_{bx} \frac{U_{ax}}{U_{Ix}} U_{cx}^* \frac{U_{cz}}{U_{Iz}} \\ &= \sum_x U_{bx} \frac{U_{ax}}{U_{Ix} U_{Iz}} \delta_{xz} \\ &= \frac{U_{bz}}{U_{Iz}} \frac{U_{az}}{U_{Iz}} = \mathsf{d}_a \mathsf{d}_b\end{aligned}$$

where we have used the fact that U is unitary. Thus we conclude

$$\mathsf{d}_a \mathsf{d}_b = \sum_c N_{ab}^c \mathsf{d}_c$$

[23] This index z must be I in well behaved modular anyon theories, but more generally in fusion rings it could be another index (Gannon [2003]).

Further Reading

The table from Rowell et al. [2009] was extended to five particle types in Bruillard et al. [2015]; moon Hong and Rowell [2010] and to six in Creamer [2019]. The modularity condition was relaxed to build a periodic table of all unitary braided theories for up to four particles in Bruillard [2016] and five in Bruillard and Ortiz-Marrero [2018]. Periodic table for theories with fermions (which are a special subset of non-modular theories) have been built in Bruillard et al. [2020] (See also Bruillard et al. [2017]).

Exercises

Exercise 17.1 Using the pivotal property

Use the pivotal property (Section 14.7.1) to demonstrate the identity shown in Fig. 17.10. You should not assume full isotopy invariance. Nor should you

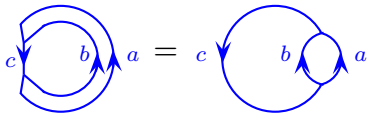


Fig. 17.10 This identity can be shown without full isotopy invariance by using the pivotal property.

assume $\epsilon = +1$ for any of the particles.

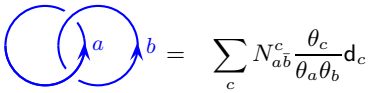
Exercise 17.2 Symmetries of S

Use isotopy of diagrams and Hermitian conjugation of diagrams to show the identities in Eq. 17.7.

Exercise 17.3 Evaluation of the S-link

(a) Use the R-matrices and Eq. 15.3 to derive the value of the matrix $\tilde{S}_{ax}$ (See Fig. 17.2) in terms of fusion multiplicities, twist factors θ_a, and the quantum dimensions d_a.

(b) From your result show that



$$= \sum_c N_{a\bar{b}}^c \frac{\theta_c}{\theta_a \theta_b} \mathsf{d}_c$$

Note that this diagram differs from S_{ab} by a factor of $Z(S^3) = 1/\mathcal{D}$.

Exercise 17.4 Product theories[Easy]

Given two anyon theories A and B with corresponding S-matrices S_A and S_B.

(a) Show that the product theory $A \times B$ has S-matrix $S_A \otimes S_B$.

(b) Show that $A \times B$ is modular if and only if both A and B are modular.

(c) Show that the central charge of the product theory is the sum of the central charges of the constituent theories. I.e.,

$$c_{A\times B} = (c_A + c_B) \bmod 8 \tag{17.23}$$

In fact, central charges strictly add in product theories. However, we have only defined the central charge mod 8 so far!

Part IV

Some Examples: Planar Diagrams and Anyon Theories

18 Some Simple Example

In this chapter we consider a few simple examples of anyon theories. Our strategy will be the same in each case. First, we decide on a set of fusion rules. From this we examine the possible planar planar diagram algebras. To a mathematician these are known as spherical tensor categories (See section 14.7.2). Once we have found the possible planar algebras we will look for possible braidings to build full anyon theories.

All of the examples given here will enjoy full isotopy invariance (with $\epsilon_a = +1$, although if there are nontrivial Frobenius-Schur indicators we will need negative d's) so we will use the notation of chapter 16.

18.1 $\mathbb{Z}_2$ Fusion Rules

Let us start with the simplest system of particles we can imagine, an identity 0 and a nontrivial particle 1. The simplest fusion rules we can have are[1]

$$1 \times 1 = 0$$

which tells us that 1 is its own antiparticle $1 = \bar{1}$ so we do not draw arrows on the corresponding line. This is known as $\mathbb{Z}_2$ fusion rules and is shown in Fig. 18.1. The corresponding fusion multiplicity matrix is $N_{11}^0 = N_{00}^0 = N_{10}^1 = N_{01}^1$ and $N_{11}^1 = N_{10}^0 = N_{01}^0 = N_{00}^1 = 0$.

[1] We have switched notations — here the vacuum is 0 not I, and the nontrivial particle is 1. I hope this does not cause confusion!

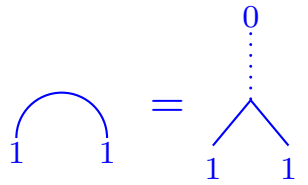


Fig. 18.1 Fusing two 1-particles to the vacuum, shown in two notations.

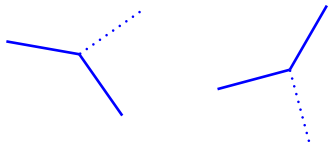


Fig. 18.2 Examples of allowed vertices for the $\mathbb{Z}_2$ fusion rules. A 1 particle (drawn solid) comes into the vertex and the 1-partice must also go out of the same vertex. The 0 particle, the identity, is drawn dotted, but it need not be drawn at all.

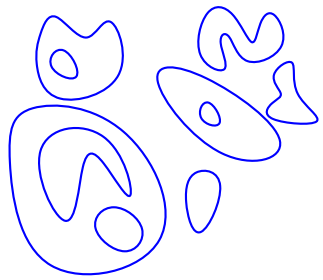


Fig. 18.3 A loop gas has $\mathbb{Z}_2$ fusion rules. The loop gas drawn here is planar — there are no over- or under-crossings.

With 0 being the identity, the only nontrivial vertices we can have with these fusion rules is where one particle 1 comes in and one particle 1 also goes out as shown in Fig. 18.2. If one does not draw the identity particle, diagrams must then be just a so-called *loop gas* as shown in Fig. 18.3. The constraint $N_{01}^0 = N_{00}^1 = N_{10}^0 = 0$ means that loops cannot end, and $N_{11}^1 = 0$ means that loops cannot intersect.

Since the largest eigenvalue of $[N_a]_b^c$ is 1, we have quantum dimensions $\mathsf{d}_1 = \mathsf{d}_0 = 1$ (See Eq. 8.9).

The $\mathbb{Z}_2$ loop gases were studied in Exercise 2.2 (where we allowed over and undercrossings in addition to just planar diagrams), and we will consider them again in section 19.1 below.

With these fusion rules, there are two sets of F-matrices that give a consistent fully isotopy invariant planar algebras. These moves two solutions correspond to $d_1 = \pm 1$, which are the only two options given $\mathsf{d}_1 = 1$.

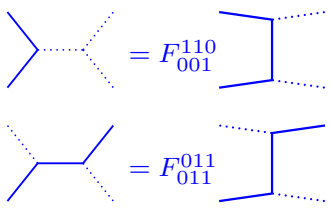


Fig. 18.4 These F-moves for the $\mathbb{Z}_2$ loop gas simply deform the path of the particles. These are known as "isotopy" moves.

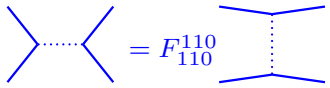


Fig. 18.5 This F-move for the $\mathbb{Z}_2$ loop gas reconnects the paths of particles. This is known as a "surgery" move.

18.1.1 $d = +1$ Loop Gas

Here we choose $d_1 = +1$ for the nontrivial particle, in which case every F which is nonzero is $+1$. (I.e., every F diagram where all vertices are consistent with the fusion rules. See Fig 16.3). In other words, $F^{bad}_{ecf} = 1$ for every case where $N_{abd} = N_{ced} = N_{bcf} = N_{aef} = 1$ and F is zero otherwise. We can write out explicitly the nonzero elements

$$F^{000}_{000} = 1/d_0 = 1 \tag{18.1}$$
$$F^{110}_{110} = 1/d_1 = 1 \tag{18.2}$$
$$F^{110}_{001} = F^{000}_{111} = 1 \tag{18.3}$$
$$F^{101}_{101} = F^{011}_{011} = 1 \tag{18.4}$$
$$F^{101}_{010} = F^{011}_{100} = 1 \tag{18.5}$$

The first two lines are required from Eq. 16.4. Eq. 18.3 is from Eq. 16.7. Eq. 18.4 and Eq. 18.5 can be derived from Eq. 18.2 and Eq. 18.3 by the tetrahedral symmetry equation Eq. 16.17. Examples of these F-moves are shown in Fig. 18.4 and 18.5.

The $d = 1$ planar loop gas turns out to be a relatively trivial diagrammatic algebra. The value of every allowed diagram is unity! (or is zero if there is anything disallowed in the diagram, such as the intersection of loops.)

We now turn to consider the possible braidings that we can impose on this planar algebra. The only nontrivial R matrix element is R^{11}_0. Using the hexagon equation 16.21 and setting $a = b = c = d = 1$ and $e = c = g = 0$ we obtain (the only allowed value of f is 0)

$$[R^{11}_0]^2 F^{110}_{110} = [F^{110}_{110}]^2 R^{10}_1 \tag{18.6}$$

The R on the right is unity, and the F's are all unity. Thus

$$[R^{11}_0]^2 = 1 \tag{18.7}$$

This limits us to two possible anyon theories for the $d = +1$ loop gas:

Bosons

We choose the $R_0^{11} = +1$ case. This gives us no phases or signs with F moves or braiding. The corresponding twist factor (via Eq. 15.1) is $\theta_a = +1$ which corresponds to bosons.

Fermions

We choose the $R_0^{11} = -1$ case. This gives us minus sign under exchange of identical partiecles. The corresponding twist factor (via Eq. 15.1) is $\theta_a = -1$ which corresponds to fermions.

For both bosons and fermions the S matrix describing the braiding is

$$S = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \tag{18.8}$$

which is not unitary, so neither of these two cases are modular[2].

[2]Since this is not a modular theory, the S matrix is not the matrix that diagonalizes the fusion rules! That matrix is Eq. 18.14.

18.1.2 $d = -1$ Loop Gas

Here we choose $d_1 = -1$ for the nontrivial particle, in which case every F which is consistent with the fusion rules is ± 1. The signs of the nonzero elements of F are given as follows

$$F_{000}^{000} = 1/d_0 = 1 \tag{18.9}$$
$$F_{110}^{110} = 1/d_1 = -1 \tag{18.10}$$
$$F_{001}^{110} = F_{111}^{000} = 1 \tag{18.11}$$
$$F_{101}^{101} = F_{011}^{011} = 1 \tag{18.12}$$
$$F_{010}^{101} = F_{100}^{011} = 1 \tag{18.13}$$

As with the $d = +1$ loop gas, the first two lines are required from Eq. 16.4. Eq. 18.3 is from Eq. 16.7. Eq. 18.4 and Eq. 18.5 can be derived from Eq. 18.2 and Eq. 18.3 by the tetrahedral symmetry equation Eq. 16.17. Note in particular how the signs work in Fig. 16.14 in the definition of the tetrahedral diagram.

It is worth looking at the two different signs that F can take. (If necessary, refer back to Fig. 16.3 for details of how the F-matrix is defined). Moves such as shown in Fig. 18.4 simply deform the path of the particle and do not incur a sign. However, the move shown in Fig. 18.5 perform "surgery" on the parths and reconnect loops and does change the sign. Such a surgery always changes the number of loops in the diagram by one. The value of any loop diagram is thus given by

$$\begin{matrix}\text{Value of } (d=-1) \\ \text{loop diagram}\end{matrix} = (-1)^{\text{number of loops}}$$

As discussed in detail in sections (See also 16.3), while this is a perfectly consistent planar diagrammatic algebra, it has non-positive definite inner products and therefore is not appropriate for describing quantum

mechanics.

However as discussed in 14.2.1 we can make make a proper unitary theory out of the $d = -1$ loop gas just by implementing rule 0. So the unitary evaluation of a diagram is

$$\begin{array}{l}\text{Value of } (d=-1)\\ \text{loop diagram}\\ \text{including rule 0}\end{array} \quad = \quad (-1)^{\text{number of loops+number of caps}}$$

For example, in Fig. 18.3 there are 10 loops and 14 caps, so the full value of the diagram is $+1$.

The nontrivial particle here has a negative Frobenius-Schur indicator. Here we have chosen to do our bookkeeping by working with a negative d, by maintaining isotopy invariance of diagrams (with $\epsilon = +1$) and implementing rule 0 to obtain a unitary theory. There are other possible ways to account for this sign, but in the end we will always have to pay the price of a minus sign for a space-time wiggle such as Fig. 14.5.

Semions

We now consider possible braidings for the $d = -1$ loop gas. As in the case $d = +1$, we can apply the hexagon to obtain Eq. 18.6. In this case, however, $F^{110}_{110} = -1$ so we obtain

$$[R^{11}_0]^2 = -1$$

Again there are two solutions $R^{11}_0 = \pm i$ corresponding to right- and left-handed semions. In either case wrapping a semion all the way around another gives -1, so the S matrix is given by

$$S = \frac{1}{\sqrt{2}} \begin{pmatrix} +1 & +1 \\ +1 & -1 \end{pmatrix} \tag{18.14}$$

This is unitary, telling us that the theory is modular.

18.2 Fibonacci Fusion Rules: The Branching Loop Gas

We now consider Fibonacci fusion rules as discussed in sections 8.2.1 and 9.1 above. Here the nontrivial fusion rule is[3]

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

Again $\tau = \bar{\tau}$ is self-dual. These fusion rules allow vertices with three τ particles (one coming from each direction as shown in Fig. 18.6) so the loop gas can have branches as shown in Fig. 18.7.

The fusion multiplicity matrix N^c_{ab} is zero if exactly one of the indices is τ and the other two are I. Otherwise $N^c_{ab} = 1$. We can establish the nonzero components of the F-matrices for these fusion rules:

[3] Here we have switched back to the notation of τ for the nontrivial particle and I for the vacuum. Using 1 and 0 is also common.

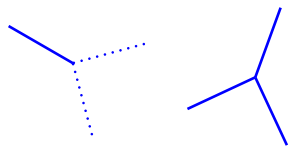


Fig. 18.6 An allowed fusion vertex (right) and a disallowed fusion vertex (left) for the Fibonacci fusion rules. The solid line is τ and the dotted line is the identity. The vertices shown in Fig. 18.2 are also allowed.

Fig. 18.7 A Fibonacci branching loop diagram allows intersections of loops, but no loop ends.

$$
\begin{aligned}
F^{III}_{III} &= 1/d_I = 1 && (18.15)\\
F^{\tau\tau I}_{\tau\tau I} &= 1/d_\tau && (18.16)\\
F^{\tau\tau I}_{II\tau} &= F^{III}_{\tau\tau\tau} = 1 && (18.17)\\
F^{\tau\tau I}_{\tau\tau\tau} &= 1/\sqrt{d_\tau} && (18.18)\\
F^{\tau\tau\tau}_{\tau\tau I} &= 1/\sqrt{d_\tau} && (18.19)\\
F^{\tau I\tau}_{\tau I\tau} &= F^{I\tau\tau}_{I\tau\tau} = 1 && (18.20)\\
F^{\tau I\tau}_{I\tau I} &= F^{I\tau\tau}_{\tau II} = 1 && (18.21)\\
F^{\tau\tau\tau}_{I\tau\tau} &= F^{\tau\tau\tau}_{\tau I\tau} = F^{I\tau\tau}_{\tau\tau\tau} = F^{\tau I\tau}_{\tau\tau\tau} = 1 && (18.22)\\
F^{\tau\tau\tau}_{\tau\tau\tau} &= -1/d_\tau && (18.23)
\end{aligned}
$$

As with the case of the $\mathbb{Z}_2$ loop gases, the first two lines are required from Eq. 16.4. Eq. 18.17 and Eq. 18.18 are from Eq. 16.7. Eq. 18.19 comes from Eq. 18.18 and Eq. 16.10. Eqs. 18.20, 18.21, and 18.22 can be derived from Eqs. 18.16, 18.17, 18.18 and 18.19 by the tetrahedral symmetry equation Eq. 16.17. Finally, Eq. 18.23 comes from the requirement that the two by two matrix $[F^{\tau\tau}_{\tau\tau}]$ is a unitary matrix (See Fig. 16.3) which we write out as[4]

$$
F^{\tau\tau}_{\tau\tau} = \begin{pmatrix} 1/d_\tau & 1/\sqrt{d_\tau} \\ 1/\sqrt{d_\tau} & -1/d_\tau \end{pmatrix} \qquad (18.24)
$$

The unitarity requirement on this matrix also gives us

$$
\frac{1}{d_\tau^2} + \frac{1}{d_\tau} = 1 \qquad (18.25)
$$

The solution to this is

$$
d_\tau = \frac{1+\sqrt{5}}{2}
$$

which matches the expected quantum dimension d_τ given in Eq. 8.2 as it must, given the considerations of chapter **??**. Eq. 18.25 also has a solution with $d_\tau < 0$. However, we cannot accept this solution because it would violate Eq. 14.4[5].

As in the case of the $\mathbb{Z}_2$ loop gases, many of the F-matrix elements correspond to simple deformations of paths (isotopy) as in Fig. 18.4. The nontrivial F-moves (corresponding to the matrix F in Eq. 18.24) are summarized in Fig. 18.8.

[4] This F-matrix matches our previous claim in Eq. 9.3. With any proposed F-matrix, one should always check that one has a valid solution of the pentagon equation Eq. 9.7 (or Eq. 16.3). See exercise 9.4 for the Fibonacci case!

[5] Since $\tau \times \tau = \tau + \ldots$ we need to have $\mathrm{sign}[d_\tau] \times \mathrm{sign}[d_\tau] = \mathrm{sign}[d_\tau]$.

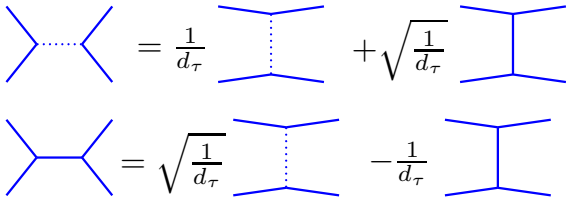


Fig. 18.8 The F-moves for the Fibonnaci branching loop gas. Note that the first line is actually the insertion of a complete set of states as in Fig. 16.8

18.2.1 Braidings for Fibonacci Anyons

To determine the possible braidings for Fibonacci fusion rules, we must solve the hexagon equation given the F matrices we just derived. This is assigned as exercise 13.1. There are two possible solutions, a right-handed solution given in Eq. 10.2, and a left-handed solutions which is the complex conjugate of the right handed solution. These are the only solutions of the hexagon equations for the Fibonacci fusion rules.

18.3 Ising Fusion Rules: A Two Species Loop Gas

As discussed in section 8.2.2 the Ising fusion rules (also known as $SU(2)_2$ fusion rules) are given by

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

with both particle types being self dual $\psi = \bar{\psi}$ and $\sigma = \bar{\sigma}$. These rules describe a loop gases with two non-vacuum particles ψ (which we draw as blue lines and loops in Fig. 18.9) and σ (which we draw red loops in Fig. 18.9). The rule of this loop gas is that one may have a vertex with two sigmas and one ψ, which appears as a blue line splitting off from a red loop.

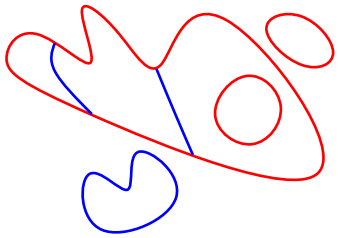


Fig. 18.9 A diagram with Ising fusion rules. Here σ is red and ψ is blue.

Looking at the first fusion rule, $\psi \times \psi = I$, we realize this rule alone,

is simply a $\mathbb{Z}_2$ fusion rule. Indeed, this tells us immediately that we have

$$\begin{aligned} & 1/d_I = 1/d_\psi = 1 \\ = \quad & F^{III}_{III} = F^{\psi\psi I}_{\psi\psi I} = F^{\psi\psi I}_{II\psi} = F^{III}_{\psi\psi\psi} = F^{\psi I\psi}_{\psi I\psi} = F^{I\psi\psi}_{I\psi\psi} = F^{\psi I\psi}_{I\psi I} = F^{I\psi\psi}_{\psi II} \end{aligned}$$

as given in Eqs 18.1-18.5. One might wonder why we do not consider $d_\psi = -1$. This is for the same reason why we could not consider negative d_τ in the Fibonacci case. Here we must have $\mathrm{sign}[d_\sigma]\mathrm{sign}[d_\sigma] = \mathrm{sign}[d_\psi]$, so we must have d_ψ positive.

Very similarly we have

$$\begin{aligned} F^{\sigma\sigma I}_{\sigma\sigma I} &= 1/d_\sigma \\ F^{\sigma\sigma I}_{II\sigma} &= F^{III}_{\sigma\sigma\sigma} = 1 \\ F^{\sigma I\sigma}_{\sigma I\sigma} &= F^{I\sigma\sigma}_{I\sigma\sigma} = 1 \\ F^{\sigma I\sigma}_{I\sigma I} &= F^{I\sigma\sigma}_{\sigma II} = 1 \end{aligned}$$

The first equation is from Eq. 16.4, and the second from Eq. 16.7. The last two are derived from the first two via the tetrahedral symmetry Eq. 16.17.

Further using Eqs. 16.7 and 16.16 we obtain

$$F^{\sigma\sigma I}_{\sigma\sigma\psi} = F^{\sigma\sigma\psi}_{\sigma\sigma I} = 1/d_\sigma \tag{18.26}$$

$$F^{\sigma\sigma I}_{\psi\psi\sigma} = F^{\psi\psi I}_{\sigma\sigma\sigma} = F^{\sigma\psi\sigma}_{\psi\sigma I} = F^{\psi\sigma\sigma}_{\sigma\psi I} = 1 \tag{18.27}$$

Enforcing unitarity on the two by two matrix $[F^{\sigma\sigma}_{\sigma\sigma}]$ we get

$$F^{\sigma\sigma\psi}_{\sigma\sigma\psi} = -1/d_\sigma \tag{18.28}$$

giving the two by two matrix the form

$$[F^{\sigma\sigma}_{\sigma\sigma}] = \begin{pmatrix} 1/d_\sigma & 1/d_\sigma \\ 1/d_\sigma & -1/d_\sigma \end{pmatrix} \tag{18.29}$$

The unitarity condition also gives us the condition that

$$d_\sigma = \pm\sqrt{2}$$

which is expected from section 8.2.2 since the fusion rules give us $\mathsf{d}_\sigma = |d_\sigma| = \sqrt{2}$. Both of these roots are viable solutions of the pentagon.

The remaining nonzero elements of F are obtained from Eq. 18.26-18.28 by using tetrahedral symmetry Eq. 16.17 to obtain

$$1 = F^{\sigma I\sigma}_{\sigma\psi\sigma} = F^{I\sigma\sigma}_{\psi\sigma\sigma} = F^{\sigma\psi\sigma}_{\sigma I\sigma} = F^{\psi\sigma\sigma}_{I\sigma\sigma} \tag{18.30}$$

$$= F^{\sigma I\sigma}_{\psi\sigma\psi} = F^{I\sigma\sigma}_{\sigma\psi\psi} = F^{\psi I\psi}_{\sigma\sigma\sigma} = F^{I\psi\psi}_{\sigma\sigma\sigma} \tag{18.31}$$

$$= F^{\psi\sigma\sigma}_{\sigma I\psi} = F^{\sigma\sigma\psi}_{I\psi\sigma} = F^{\sigma\sigma\psi}_{\psi I\sigma} = F^{\sigma\psi\sigma}_{I\sigma\psi} \tag{18.32}$$

$$-1 = F^{\sigma\psi\sigma}_{\sigma\psi\sigma} = F^{\psi\sigma\sigma}_{\psi\sigma\sigma} \tag{18.33}$$

The nontrivial F-moves corresponding to the matrix Eq. 18.29 are

shown in Fig. 18.10.

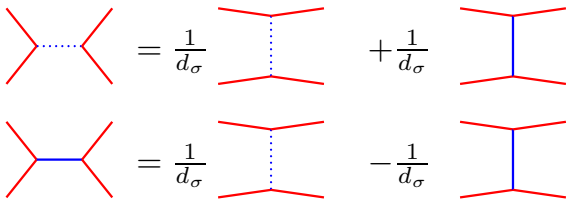


Fig. 18.10 The nontrivial F-moves for the Ising fusion rules. Note that the first line is actually the insertion of a complete set of states as in Fig. 16.8

18.3.1 Braidings For Ising Fusion Rules

[6]Since the F matrices are real, Eq. 13.3 and 16.22 are equivalent to Eq. 13.2 and 16.21.

The most straightforward way to find all the possible braidings for the Ising fusion rules is to explicitly solve the hexagon equations 13.2 or equivalently Eq. 16.21[6] . We here outline how we proceed (Exercise 18.1 asks you to work out the details!). For each possible settings of the variables in the hexagon equation, we derive a different identity. For each of the following cases, the F-matrices are simple scalars (1 and -1 only) so we derive

$$a=\psi,\ b=c=\sigma,\ d=I \quad \Rightarrow \quad R_\sigma^{\sigma\psi}R_\psi^{\sigma\sigma}=R_I^{\sigma\sigma} \tag{18.34}$$

$$b=\psi,\ a=c=\sigma,\ d=I \quad \Rightarrow \quad R_\sigma^{\psi\sigma}R_\psi^{\sigma\sigma}=R_I^{\sigma\sigma} \tag{18.35}$$

$$a=b=\sigma,\ c=\psi,\ d=I \quad \Rightarrow \quad [R_\sigma^{\psi\sigma}]^2=R_I^{\psi\psi} \tag{18.36}$$

$$a=b=\sigma,\ c=d=\psi \quad \Rightarrow \quad [R_\sigma^{\psi\sigma}]^2=-1 \tag{18.37}$$

For the following case one uses the two-by-two F matrix meaning we are working with a two dimensional vector space and the hexagon gives us two identities

$$a=b=c=d=\sigma,\ e=I \Rightarrow \begin{cases} [R_I^{\sigma\sigma}]^2=\frac{1}{d_\sigma}(1+R_\sigma^{\psi\sigma}) \\ R_I^{\sigma\sigma}R_\psi^{\sigma\sigma}=\frac{1}{d_\sigma}(1-R_\sigma^{\psi\sigma}) \end{cases} \tag{18.38}$$

These equations are enough to pin down all of the possible solutions for the R-matrix. From Eq. 18.36 and 18.37 we obtain

$$R_I^{\psi\psi}=-1$$

which also implies $\theta_\psi=-1$ from Eq. 15.2. Note that since ψ is a $\mathbb{Z}_2$ field with d_ψ comparing to our above discussion of the $\mathbb{Z}_2$ fusion rules we already knew that we had to have ψ be either a fermion or a boson (we could re-establish this by looking at the hexagon with only ψ and I fields). The hexagon including the σ field now establishes ψ to be a fermion!

From Eqs. 18.34 and 18.35 and 18.37 we establish

$$R_\sigma^{\sigma\psi} = R_\sigma^{\psi\sigma} = \pm i$$

This sign is an additional free choice we can make (in addition to the choice of $d_\sigma = \pm\sqrt{2}$). To keep these independent choices straight we will use the notation

$$d_\sigma = \overset{1}{\pm}\sqrt{2} \tag{18.39}$$

$$R_\sigma^{\psi\sigma} = \overset{2}{\pm} i \tag{18.40}$$

We now plug in our choices for d_σ and $R_\sigma^{\psi\sigma}$ into the first of Eq. 18.38 to solve for $R_I^{\sigma\sigma}$. This gives yet another independent choice of sign for a square root which we label with $\overset{3}{\pm}$. We thus obtain

$$R_I^{\sigma\sigma} = \exp\left[2\pi i\left(\frac{3}{8}\right)\right]\exp\left[2\pi i\left(\overset{3}{\pm}\frac{1}{4}\overset{1}{\mp}\frac{1}{8}\overset{2}{\pm}\frac{1}{16}\right)\right]$$

Note now that

$$\theta_\sigma^* = R_I^{\sigma\sigma}\text{sign}[d_\sigma] = \exp\left[2\pi i\left(\frac{5}{8}\right)\right]\exp\left[2\pi i\left(\overset{3}{\pm}\frac{1}{4}\overset{1}{\mp}\frac{3}{8}\overset{2}{\pm}\frac{1}{16}\right)\right] \tag{18.41}$$

from which we see there are a total of eight possible choices, and they are all of the possible odd-sixteenth roots of unity which is what we would predict from the fusion rules given Eq. 17.18.

For the record, from Eq. 18.34, we also have

$$R_\psi^{\sigma\sigma} = \overset{2}{\mp} iR_I^{\sigma\sigma} = \exp\left[2\pi i\left(\frac{3}{8}\right)\right]\exp\left[2\pi i\left(\overset{3}{\pm}\frac{1}{4}\overset{1}{\mp}\frac{1}{8}\overset{2}{\mp}\frac{3}{16}\right)\right]$$

We should also check that none of the other hexagon relations are violated for any of these eight solutions (We could for example, evaluate the hexagon with $a = d = \psi$, $b = c = \sigma$ or any of the other possible combinations). Remarkably, perhaps, all eight solutions solve all the hexagon relations with no violations (see exercise 18.1).

The eight possible solutions all have the same S-matrix (see exercise 18.2)

$$S = \frac{1}{2}\begin{pmatrix} 1 & \sqrt{2} & 1 \\ \sqrt{2} & 0 & -\sqrt{2} \\ 1 & -\sqrt{2} & 1 \end{pmatrix} \tag{18.42}$$

The two possibilities

$$\begin{array}{lll} \overset{1}{\pm} = + & \overset{2}{\pm} = - & \overset{3}{\pm} = - \\ \overset{1}{\pm} = + & \overset{2}{\pm} = + & \overset{3}{\pm} = - \end{array}$$

gives the Ising TQFT and its conjugate respectively. These cases have d_σ chosen positive.

The two possibilities

$$\begin{array}{ccc} \overset{1}{\pm}= - & \overset{2}{\pm}= + & \overset{3}{\pm}= - \\ \overset{1}{\pm}= - & \overset{2}{\pm}= - & \overset{3}{\pm}= + \end{array}$$

correspond to $SU(2)_2$ Chern-Simons theory and its conjugate respectively. Both of these cases have d_σ negative.

Further Reading

Exercises

Exercise 18.1 Using the Hexagon for Ising Fusion Rules
Use the hexagon relations to derive Eqns. 18.34-18.38. Confirm that the eight solutions we find give no violations of any hexagon relations.

Exercise 18.2 S-matrix for Ising Fusion Rules
Explicitly derive the S-matrix for all eight solutions of the hexagon equation for the Ising fusion rules and confirm that they all give Eq. 18.42. Thus confirm that all eight solutions are modular. Hint: It might be easiest to use Eq. 17.16.

Exercise 18.3 Frobenius-Schur Indicator for Ising Fusion Rules
Use Eq. 17.17 to calculate the Frobenius-Schur indicator for the σ particle in each of the eight possible solutions of the hexagon equation for the Ising fusion rules. Show that the Frobenius-Schur indicator is negative exactly when d_σ is negative.

Exercise 18.4 Evaluating Diagrams I
Show that evaluation of the diagram in Fig. 18.7 gives $-d_\tau^{9/2}$.

Exercise 18.5 Evaluating Diagrams II
Show that evaluation of the diagram in Fig. 18.9 gives $d_\psi^2 d_\sigma^3 \kappa_\sigma$.

Exercise 18.6 Deriving an F-matrix
(Easy) Consider a theory containing three particle types, I, A, B where I is the identity. Let the nontrivial fusion rules be given by

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

Let us assume we have a theory with full isotopy invariance and full tetrahedral symmetry. There is only one set of F matrices for these fusion rules. Find these F-matrices and convince yourself that it satisfies the pentagon equation.

Exercise 18.7 Deriving an F-matrix
(Hard) Consider a theory containing three particle types, I, S, V where I is the identity. Let the nontrivial fusion rules be given by

$$S \times S = I$$

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

Let us assume we have a theory with full isotopy invariance and full tetrahedral symmetry. There is only one set of F matrices for these fusion rules. Find these F-matrices. (Hint: applying the techniques of this chapter will give the correct solution as well as a spurious solution which you can eliminate by showing it does not satisfy the pentagon equation.) Show that the hexagon equaution is solved with $R_c^{ab} = 1$ for all a, b, c.

Temperly-Lieb Algebra and Jones-Kauffman Anyons

19

Let us look back at the Kauffman bracket invariant that we introduced in chapter 2. In the current chapter we want to make use of these rules and determine some of the properties of the corresponding anyons in the langauge we have been developing since chapter 8. Our strategy will be to first consider a planar diagram algebra in detail before considering braiding properties in section 19.5 below.

So we start by considering a planar version of the Kauffman bracket. I.e., we only consider diagrams with with no over- and under-crossings. Our diagrams are isotopically invariant in the plane and the only additional rule then is that the a loop is given a value d as shown in Fig. 19.1. As compared to the diagrammatic algebra we have constructed over the last few chapters (roughly starting in chapter 8, and continuing through chapter 16), one things that was missing in the discussion of the Kauffman bracket invariant is the idea of multiple particle types and fusion rules. In this chapter we will try to construct particle types, fusion rules, and F-matrices given only the rule 19.1 as a starting point. The planar algebra of loops that we will construct is known as the Temperly-Lieb algebra. (When we reintroduce braiding to our theory the resulting theory is called called Jones-Kauffman, or Temperley-Lieb-Jones-Kauffman.)

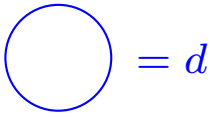


Fig. 19.1 The loop rule for the Kauffman bracket invariant and the Temperly-Lieb algebra.

Let us start by thinking a bit about what kind of particle types we already have in our theory. Certainly we have the simple string[1] which we will call "1"; and we always have a vacuum particles, which we will call "0". Now we would like to ask whether we can fuse two of these 1-strings together to make another particle.

[1] It is admittedly confusing that 1 is not 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 nontrivial.

Several things are immediately obvious. First consider the fact that two 1-particles can fuse to the vacuum, or in other words, a 1-string can go up and then turn down, as shown in Fig. 19.2. This tells us immediately that

$$1 = \bar{1}.$$

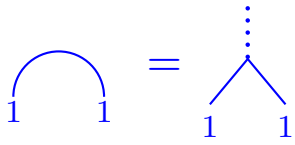


Fig. 19.2 Fusing two 1-particles to the vacuum

The fact that 1 is its own antiparticle is why we do not draw arrows on the 1-string. For simplicity, if a string is not labeled we will assume it is a 1-string. Given that loop of 1-string is assigned the value d, we identify this d_1 (which is often called the quantum dimension, although we have been reserving the words "quantum dimension" for $\mathsf{d}_1 = |d_1|$).

We might also consider the possibility that two of these 1-particles can fuse to something besides the vacuum, in a way similar to that shown in Fig. 19.3. This is a good idea, but it isn't yet quite right. If

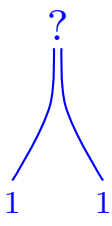


Fig. 19.3 Attempting to Fuse two 1-particles to something different from the vacuum

the two strings fuse to some object besides the vacuum 0, we have to make sure that this new object is appropriately "orthogonal" to 0. This orthogonality must be in the sense of the locality, or no-transmutation rule (see Fig. 8.7): a particle type must not be able to spontaneously turn into another particle type (without fusing with some other particle or splitting). In Fig. 19.3 it looks like the two strings brought together could just fuse together to form the vacuum as in Fig. 19.2, and this would then turn the collection of two strings into the vacuum. To prevent such transmutation, we will work with operators known as Jones-Wenzl projectors.

19.1 Jones-Wenzl Projectors

Fig. 19.4 A cup (left) and a cap (right).

The general definition of a projector is an operator P such that $P^2 = P$. This means that P has eigenvalues 0 and 1. Let us think of a string diagram as an operator that takes as an input strings coming from the bottom of the page, and gives as an output strings going towards the top of the page (compare Fig. 12.2). Now consider a set of n-strings traveling together in the same direction (in what is often called a *cable*). The Jones-Wenzl projection operator P_n operates on a set of n such strings — it takes n-strings in and gives n-strings out — and it is defined such that attaching a cup or a cap to the bottom or top of the operator gives a zero result (See Fig. 19.4). The n-particle Jones-Wenzl projector P_n acting on a cable of n-strings should be interpreted as the n^{th} particle species.

The purpose of the Jones-Wenzl projector is to fix the problem we discovered with Fig. 19.3. That is, if a cable of two strings forms a nontrivial particle (the particle we will call 2), we should not be able to put a cap on the top of these two strings and transmute the 2-particle to the vacuum. I.e., adding a cap should make the entire diagram vanish, and this is the property we are looking for in the 2-string Jones-Wenzl projector.

Let us now try to construct the 2-string Jones-Wenzl projector P_2 out of two incoming 1-particles[2] (two elementary strings). To do this we first construct a different projector $\bar{P}_2$ that forces the two incoming particles to fuse to the vacuum[3] as shown in Fig. 19.5.

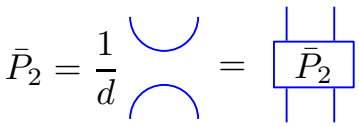


Fig. 19.5 The projector of two strings to the vacuum $\bar{P}_2$. This figure should be thought of as an operator that takes as an input two strings coming in from the bottom, and gives as an output two strings going out the top. Sometimes the operator is represented as a labeled box as shown on the right.

To establish that this $\bar{P}_2$ operator is a projector we need to check that $[\bar{P}_2]^2 = \bar{P}_2$. To apply the $\bar{P}_2$ operator twice we connect the two strings coming out the top of the first operator to two strings coming in the

[2]The Jones-Wenzl projector, if you want to defines one, for a single string is the trivial operator. I.e., one string comes in and the same string comes out unchanged.

[3]The estute reader will notice that a particle "turning around" as in Fig. 19.2 is not quite the same as projecting to the 0 particle, due to the prefactor $1/d$. We will return to this issue in section 19.3 below.

bottom of the second operator. As shown in Fig. 19.6, using the fact that a loop gets value d we see that $[\bar{P}_2]^2 = \bar{P}_2$ meaning that $\bar{P}_2$ is indeed a projector.

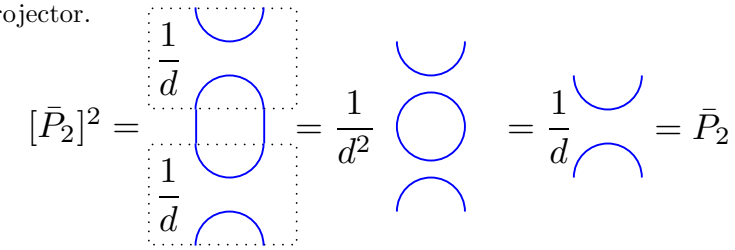


Fig. 19.6 Checking that $[\bar{P}_2]^2 = \bar{P}_2$. In the second step we have used the fact that a loop gets the value d.

The Jones-Wenzl projector P_2 for two strings is the complement of the operator $\bar{P}_2$ we just found, meaning $P_2 = I - \bar{P}_2$ where I is the identity operator, or just two parallel strings. Diagrammatically we have Fig. 19.7. Since the $\bar{P}_2$ operator projects the two strings onto the vacuum, the P_2 operator projects the two strings to a different orthogonal particle type which we call 2.

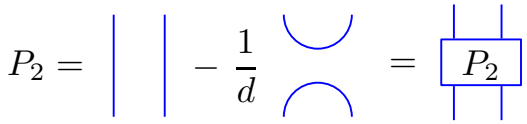


Fig. 19.7 The projector of two strings to the nontrivial particle made of two strings $P_2 = I - \bar{P}_2$. Sometimes this projector is drawn as a labeled box, as on the right.

We can algebraically check that P_2 is indeed a projector

$$P_2^2 = (I - \bar{P}_2)(I - \bar{P}_2) = I - 2\bar{P}_2 + \bar{P}_2^2 = I - \bar{P}_2 = P_2$$

and also we can check that P_2 is orthogonal to $\bar{P}_2$, by

$$\bar{P}_2 P_2 = \bar{P}_2(I - \bar{P}_2) = \bar{P}_2 - \bar{P}_2^2 = 0$$

and similarly $P_2\bar{P}_2 = 0$.

Often it is convenient to draw these projection operators as a labeled box, as shown on the right of Figs. 19.5 and 19.7. Sometimes instead of drawing two lines with a projector $\bar{P}_2$ or P_2 inserted, we simply draw a single line with a label, 0 or 2 respectively as in the right of Fig. 19.10 or the left of Fig. 19.8.

It is useful to calculate the value of the of the 2-string loop[4]. This is shown in Fig. 19.8.

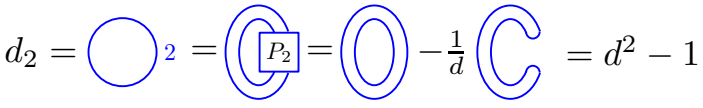


Fig. 19.8 Evaluating the 2-string loop.

[4] In many references d_2 is called Δ_2 (and similarly d_n is called Δ_n). We will stick with d to fit with the notation in the rest of this book.

For $d = \pm 1$:

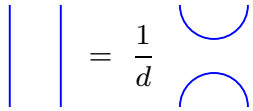


Fig. 19.9 Two cases where the Kauffman bracket invariant rules become very simple. If you have not convinced yourself of these rules, try to do so! (See exercise 2.2). Note that $d = 1$ occurs for bosons or fermions and $d = -1$ occurs for semions.

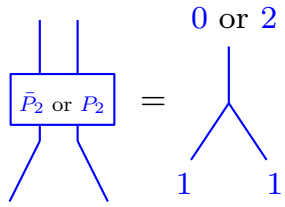


Fig. 19.10 Two possible fusions of two 1-strands, drawn in two different notations. A single line labeled 2 is interpreted as two 1-strands traveling together with a P_2 operator inserted. The label 0 means the two strands fuse to the vacuum as in Fig. 19.2.

Abelian Case

In the case where $d = \pm 1$ it is easy to prove (see Exercise 2.2 and ***) that two horizontal strings equals d times two vertical strings as shown in Fig. 19.9. In this case, notice that the projector $P_2 = 0$ since the two terms in the projector in Fig. 19.7 are equal with opposite signs. Correspondingly note that for $d = \pm 1$ (and only for these values), the value of the 2-string loop is $d_2 = 0$ as shown in Fig. 19.8, meaning that no such 2-particle exists. Thus the only possible outcome of fusion of two 1-strings is the vacuum as shown in Fig. 19.2. Thus the entire fusion rules of these theories are

$$1 \times 1 = 0$$

where again 0 is the identity or vacuum. These abelian fusion rules result in abelian braiding statistics.

These two possible cases here obviously correspond to the $d = \pm 1$ loop gases that we studied in sections 18.1.1 and 18.1.2 above. When braidings are considered we obtain bosons or fermions for $d = 1$ and left or right handed semions $(SU(2)_1)$ for $d = -1$. Since this has been discussed in depth in section 18.1 we do not elaborate further here.

Two Strands in the General Case

For values of d not equal to ± 1, the projector P_2 does not vanish. This means that two 1-strands can fuse to either 0 or 2 as shown in Fig. 19.10. 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 $\bar{P}_2 + P_2 = I$, which means these two particle types form a complete set ($\bar{P}_2$ projects the two particles to the vacuum, and P_2 projects to the 2-particle type).

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. The most general three legged operator we can construct is of the form in Fig. 19.11.

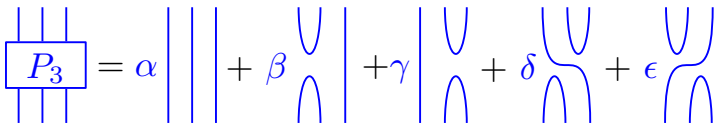


Fig. 19.11 The form of the most general three legged operator we can construct. Where $\alpha, \beta, \gamma, \delta, \epsilon$ are arbitrary constants.

We would like to find the three-string operator which is a projector. So we should enforce $P_3^2 = P_3$. However, there are other things we want to enforce as well. Since 0 is the identity, we want $0 \times 1 = 1$ which means we should not be able to fuse $\bar{P}_2$ (the projector of two strings onto the vacuum) with a single strand to get P_3. Diagrammatically this means we must insist on relations like Fig. 19.12.

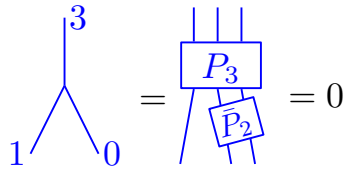


Fig. 19.12 Insisting that 0×1 does not give 3

This and analogous constraints allow us to insist on the conditions shown in Fig. 19.13.

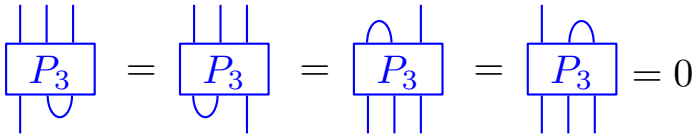


Fig. 19.13 Four conditions that come from the fusion condition shown in Fig. 19.12.

However, we should alllow fusions of the form $1 \times 2 = 3$ as shown in Fig. 19.14. Enforcing the condition in Fig. 19.13, along with $P_3^2 = P_3$ gives the form of P_3 shown in Fig. 19.11 with the results that (see Exercise 19.1)

$$\begin{aligned} \alpha &= 1 \\ \beta = \gamma &= -\frac{d}{d^2-1} \\ \delta = \epsilon &= \frac{1}{d^2-1} \end{aligned}$$

We can do a short calculation in the spirit of Fig. 19.8 to obtain the value of a loop of 3-string[4], giving the result (See exercise 19.1) shown in Fig. 19.15.

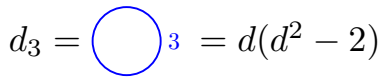


Fig. 19.15 Evaluating the 3-string loop.

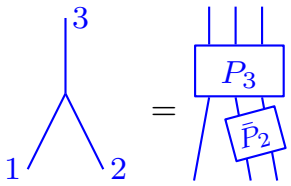


Fig. 19.14 We allow $1 \times 2 = 3$

Ising Anyons

In the case where $d = \pm\sqrt{2}$ (and only in these cases) the three string loop has $d_3 = 0$ meaning that there is no 3-string particle. Equivalently it is possible to show that P_3 vanishes when evaluated in any diagram

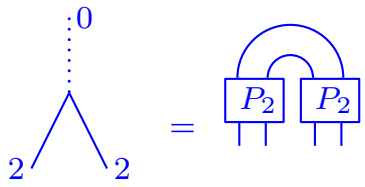


Fig. 19.16 $2 \times 2 = 0$.

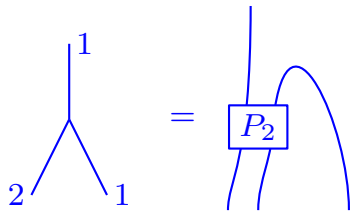


Fig. 19.17 $2 \times 1 = 1$. We recognize this as the fusion $1 \times 1 = 2$ from Fig. 19.10 just turned on its side.

For $d = \pm\sqrt{2}$:

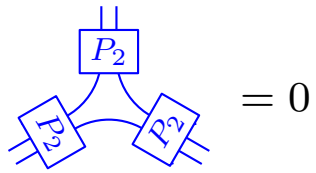


Fig. 19.18 For $d = \pm\sqrt{2}$ we have 2×2 not fusing to 2.

(See exercise 19.1). It is similarly possible to show that $P_4 = 0$ and so forth. Thus, for the case of $d = \pm\sqrt{2}$ there are only three particle types 0, 1, and 2. In addition to the fusions we have already determined, we have $2 \times 2 = 0$ as shown in Fig. 19.16 and $2 \times 1 = 1$ as shown in Fig. 19.17. (Note that showing $2 \notin 2 \times 2$ requires another explicit calculation, not shown here! See exercise 19.1)

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 sections 8.2.2 and 18.3) where $1 = \sigma$ and $2 = \psi$ and 0 is the vacuum I

Recall in our discussion of Ising anyons in sections 8.2.2 and 18.3. There we found that $d_\sigma = \pm\sqrt{2}$ and $d_\psi = 1$. This indeed agrees with the present discussion: We obtain Ising fusion rules for $d_1 = d_\sigma = \pm\sqrt{2}$ and evaluating using Fig. 19.8, we also have $d_\psi = d_2 = 1$. Thus our string algebra recovers details of the Ising fusion algebra.

19.2 General Values of d

The generalization of the above discussions for $d = \pm 1$ and $d = \pm\sqrt{2}$ is fairly straightforward. One can generally show the following properties (See Kauffman and Lins [1994] and exercise 19.2). First, the Jones-Wenzl projector for $n + 1$ strands can always be written in terms of the projector for n strands as shown in Fig. 19.19 (See exercise 19.2)

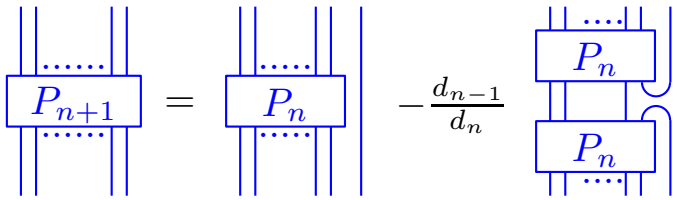


Fig. 19.19 Recursion Relation For Jones-Wenzl Projectors

Note in particular that if P_n vanishes, we can conclude that P_m vanishes for all $m > n$ as well. We define d_n of particle type n by connecting n strings coming from the bottom of projector P_n to those coming from the top as shown in Fig. 19.20.

Using the recursion shown in Fig. 19.19 and the definition of d_n in Fig. 19.20 we obtain the recursion relation (you can do this in your head!)

$$d_{n+1} = d\, d_n - d_{n-1} \tag{19.1}$$

where we define $d_{-1} \equiv 0$ and $d_0 = 1$ and hence $d_1 = d$. This recursion has the general solution

$$d_n = U_n(d/2) \tag{19.2}$$

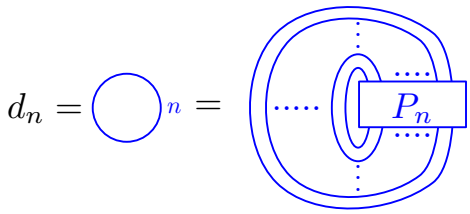


Fig. 19.20 Evaluating the quantum dimension of the n-string particle. We connect the n strings coming from the top of the projector P_n to those coming from the bottom. Often this quantity is notated as Δ_n.

where U_n is the n^{th} Chebyshev polynomial of the second kind. These are defined by (See exercise 19.2)

$$U_n(\cos\theta)\sin\theta = \sin[(n+1)\theta] \tag{19.3}$$

A theory has a finite number of particle types if $d_n = 0$ for some n (Such that P_n vanishes for all $p \geq n$). This situation occurs precisely when (See exercise 19.2)

$$d = 2\cos\left(\frac{k\pi}{n+1}\right) \tag{19.4}$$

for[5] some $k \in 1, \ldots, n$. For values of d that are not of this form, one can construct an infinite number of orthogonal particle types (n-strand projectors with different values of n), which indicates a badly behaved theory. (I.e., the algebra never "closes").

[5] For odd n the $k = (n+1)/2$ case corresponds to the unphysical case of $d = 0$.

Once one constructs the appropriate n-strand projectors, the general vertex between three different particle types can be constructed analogous to that shown in Fig. 19.21. Consider a vertex between particle types (a, b, c) as in with $a, b, c, \geq 0$ as in Fig. 19.22. The number of strings going between the projectors (as in the fight of Fig. 19.21) is given by

$$m = (a+b-c)/2 = \text{strings between } a \text{ and } b \tag{19.5}$$
$$n = (a+c-b)/2 = \text{strings between } a \text{ and } c \tag{19.6}$$
$$p = (b+c-a)/2 = \text{strings between } b \text{ and } c \tag{19.7}$$

these quantities must be non-negative, and we must have all of these

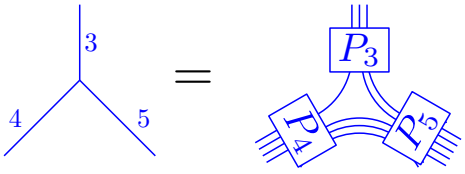


Fig. 19.21 The general vertex in the Temperly-Lieb algebra. Here the vertex is shown for 4 and 5 fusing to 3.

quantities integer, which is assured if

$$(a+b+c) \text{ is even} \tag{19.8}$$

Note that a, b or c are allowed to have the value 0, meaning no strings come out that edge. These variables are also allowed to have the value 1 meaning a single string comes out the edge (and no projector is needed, see note 2.)

One can show that a vertex between particle types (a, b, c) can be nonzero only if further if the projector

$$P_{(a+b+c)/2} \quad \text{is nonzero} \tag{19.9}$$

This final conditon is nontrivial and we will not prove it in all generality here (See for example, Kauffman and Lins [1994], for a proof). However, Fig. 19.18 is an example of this condition: When $d = \pm\sqrt{2}$, we've shown that P_3 vanishes and this implies the vertex $(2, 2, 2)$ must also vanish.

The conditions we have just described for a vertex (m, n, p non-negative integers and $P_{(a+b+c)/2}$ nonzero) gives us the fusion relations for the theory which are given by

$$a \times b = |a-b|,\ |a-b|+2,\ \ldots,\ \min(a+b, 2k-a-b)$$

where k is the largest integer such that P_k is non-zero.[6]

With this definition of a vertex we can evaluate any planar diagram. A particularly useful diagram is the version of the Theta diagram shown in Fig. 19.23. The value of this diagram can be derived generally and is given by

$$\Delta(a,b,c) = (d_{(a+b+c)/2})!\ \frac{(d_{n-1})!\,(d_{m-1})!\,(d_{p-1})!}{(d_{a-1})!\,(d_{b-1})!(d_{c-1})!} \tag{19.11}$$

where we have defined

$$(d_n)! \equiv d_n d_{n-1} d_{n-2} \ldots d_2 d_1$$

with $d_1 = d$ and $d_0 = d_{-1} = 1$. From Eq. 19.11 we see that $\Delta(a, b, c)$ is symmetric in exchanging any of its arguments. Further we see that the quantity vanishes when $d_{(a+b+c)/2}$ vanishes which agrees with the condition Eq. 19.9.

While the most general derivation of Eq. 19.11 is somewhat complicated (See Kauffman and Lins [1994]), it is easy enough to confirm it is correct for a few examples (See exercise 19.3)

The value, Eq. 19.11, of the Theta diagram does not match what we would have expected given the rules in chapter 16. Comparing to Fig. 16.13 we would have expected the Theta diagram in Fig. 19.23 to have a value $\sqrt{d_a d_b d_c}$ which in general it does not here. We will now fix this problem.

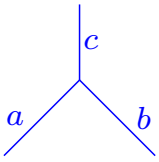


Fig. 19.22 A general vertex between particle types (a, b, c) with $a, b, c \geq 0$

[6] This is entirely equivalent to the general $SU(2)_k$ Chern-Simons fusion rules where particles j take integer and half-integer values and

$$j_1 \times j_2 = |j_1 - j_2|,\ |j_1 - j_2| + 1,\ \ldots,\ \min(j_1 + j_2, k - j_1 - j_2) \tag{19.10}$$

where we have made the identification that a in the Temperly-Lieb-Jones-Kauffman theory is $2j$. Note further that in the case where k is infinitely large (so that the final term in the series on the right of Eq. 19.10 is always j_1+j_2), these fusion rules match the angular momentum addition rules of regular $SU(2)$.

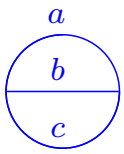
$\Delta(a,b,c) =$ 

Fig. 19.23 The Theta diagram in the Temperly-Lieb-Jones-Kauffman theory.

19.3 Unitarization

The diagrammatic algebra we have constructed so far in this chapter is a perfectly self-consistent algebra (See Kauffman and Lins [1994] for a large amount of detail of this algebra). However , this algebra does not fit the rules we have establshed in prior chapters. In section 19.2 we just found that the value of the Theta diagram does not match the expectation from chapter 16. If we tried to work out further details of the diagrammatic algebra, we would find other failures as well — for example, we would find the F-matrices to be non-unitary! Fortunately, it is not hard to modify the theory a small amount so that it fits within our existing framework from chapter 16.

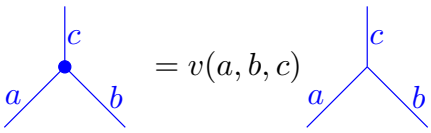


Fig. 19.24 A renormalized vertex between particle types (a, b, c) with $a, b, c \geq 0$ marked with a blue dot on the left is defined in terms of the original vertex on the right. We assume here that the vertex on the right, defined analogous to Fig. 19.21 is nonzero.

Let us define a new vertex which is a constant multiple of the old vertex as shown in Fig. 19.24. We define the rescaling factor as

$$v(a, b, c) = \sqrt{\frac{\sqrt{d_a d_b d_c}}{\Delta(a, b, c)}}$$

such that the value of the Theta diagram in Fig. 19.25 is now $\Theta(a, b, c) = \sqrt{d_a d_b d_c}$ as we expect from Fig. 16.13. It turns out that this simple modification is sufficient to make the theory fit into the framework developed in chapter 16.

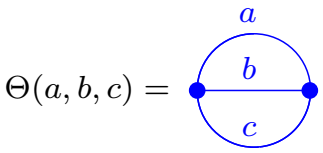


Fig. 19.25 The Theta diagram with renormalized vertices.

19.4 F-matrices

We can now determine the F-matrices directly from the graphical algebra. As a simple example, consider the F-matrices $F^{11\alpha}_{11\beta}$ (which we abbreviate as F^{α}_{β}) as shown in Fig. 19.26. Note that for this equation we use renormalized vertices as defined in Eq. 19.24 and notated by dots on the vertices.

This F-matrix equation is that of Fig. 16.3 for four incoming 1-string particles. The F matrix is nontrivial since there is more than one fusion channel when we fuse the 1's together: $1 \times 1 = 0 + 2$, so long as $d \neq \pm 1$ (in which case the 2-string particle vanishes). We can now rewrite the F-matrix equation in terms of string diagrams as in Fig. 19.27. Note that in Fig. 19.27, the prefactors of $d/\sqrt{d_2}$ come from the vertex renormalization factors $v(1, 1, 2)^2$, and the quantities in brackets are P_2 projectors which force the two strings to fuse to the 2-particle.

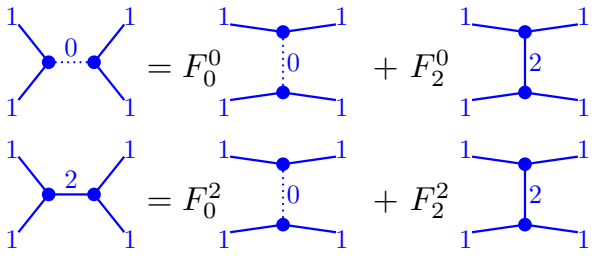


Fig. 19.26 The F-matrix in the Temperly-Lieb-Jones-Kauffman theory is unitary when use renormalized vertices, indicated by dots. Here we have abbreviated $F^{11\alpha}_{11\beta}$ as F^{α}_{β} for brevity.

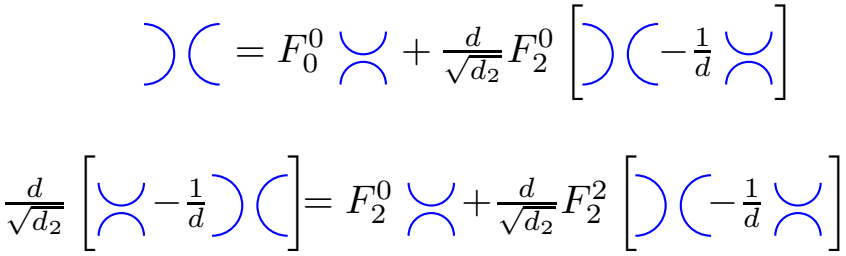


Fig. 19.27 Explicitly writing out the F-matrix equations of Fig. 19.26. The prefactors terms in brackets are P_2 projectors. The prefactors d_2 is from the vertex renormalization factors $v(1,1,2)^2 = d^2/d_2$. (The other renormalization factor $v(1,1,0) = 1$).

We then match up terms on the right and left of the graphical equations in Fig. 19.27. In the first line we see that the diagram on the left is topologically like the first term in the brackets on the right, so we have $F_2^0 = \sqrt{d_2}/d$. Similarly the first term on the right is topologically the same as the second term in the brackets, so $F_0^0 = 1/d$. Then in the second line the second term in brackets on the left is topologically the same as the first term in brackets on the right, so we have $F_2^2 = -1/d$. 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 topologically the same, so we have $d/\sqrt{d_2} = F_2^0 - (1/\sqrt{d_2})F_2^2$ or $F_2^0 = (1/d)(d^2-1)/\sqrt{d_2}$. Finally using $d_2 = (d^2-1)$ (See Fig. 19.8) we obtain the full form of the F-matrix (and returning the 11 superscripts and subscripts which we have suppressed)

$$[F_{11}^{11}] = \begin{pmatrix} \frac{1}{d} & \frac{\sqrt{d^2-1}}{d} \\ \frac{\sqrt{d^2-1}}{d} & -\frac{1}{d} \end{pmatrix} \tag{19.12}$$

Note that this matrix is properly unitary for any value of d. For $d = \pm\sqrt{2}$ the matrix matches our expectation for the Ising fusion rules given in Eq. 18.29.

With similar diagrammatic calculations, we can work out the F-

matrices for any incoming and outgoing n-string particles. Detailed calculations are given in Kauffman and Lins [1994]. However, note that the results given there are nonunitary expressions due to the use of unrernormalized vertices.

19.5 Twisting and Braiding

So far we have not yet used the braiding rules of the Kauffman bracket invariant, we have only used the loop rule and we have only considered planar diagrams. We finally can reintroduce the braiding rules for the Kauffman invariant for evaluating crossings as in Fig. 2.3 and thus we are now considering a full anyon theory. As shown in Fig. 2.6, comparing to Fig. 15.3 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 (See exercise 19.5), as well as the R-matrices. Just to do a simple example, let us evaluate R_{11}^2 as shown in Fig. 19.28.

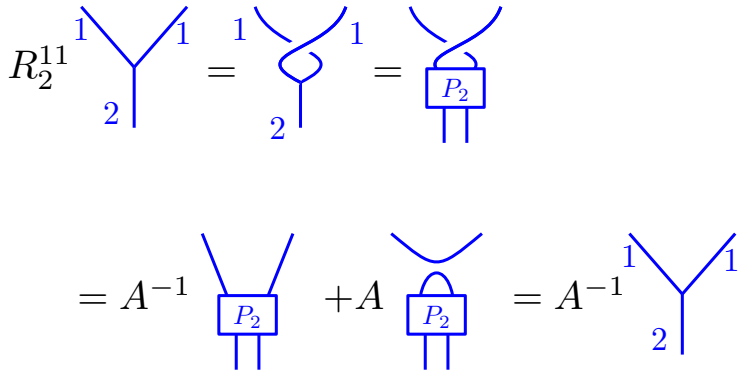


Fig. 19.28 Evaluation of $R_2^{11} = A^{-1}$ using the Kauffman bracket invariant. In going from the first line to the second we invoke the bracket rules Fig. 2.3. The last step invokes the fact that P_2 is a projector orthogonal to the turn-around thus killing the term with coefficient A.

We have already seen that in order to have a well-behaved theory with a finite number of particle types, d must take some very special values (Eq. 19.4). Also recall from Fig. 2.3 that d is related to A via

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

This means that we must have (with $k = 1, \ldots, n$)[5]

$$-A^2 = \exp\left(\pm i\pi k/(n+1)\right)$$

or

$$A = \exp\left(\overset{1}{\pm} 2\pi i \left[\frac{1}{4} \overset{2}{\pm} \frac{k}{4(n+1)}\right]\right) \tag{19.13}$$

where we have labeled the two different $\pm$ with two different superscripts to keep track of them. Note that the first $\pm$ (with the superscript 1), just changes the overall chirality of the theory.

Let us take, for example, the $n = 3$ case, meaning P_3 vanishes and we have Ising fusion rules as discussed in section 19.1. There are two possible[5] values of k given by $k = 1$ corresponding to $d = \sqrt{2}$ and $k = 3$ corresponding to $d = -\sqrt{2}$. Each of these has four possible choices of the signs in Eq. 19.13, thus resulting in eight possible anyon theories. This matches the eight theories with Ising fusion rules that we found in section 18.3 above.

Just for completeness, let us determine the twist factors for the σ particle for each of these anyon theories. We have

$$\theta_a^* = -A^3 = \exp\left(\overset{1}{\pm} 2\pi i \left[-\frac{1}{4} \overset{2}{\pm} \frac{k}{16}\right]\right)$$

with $k = 1$ or 3. This gives all the possible odd-sixteenth roots of unity as in Eq. 18.41.

Further Reading

- Louis Kauffman, *Knots and Physics*, World Scientific, (2001), 3ed. Kauffman [2001]
- 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). Kauffman and Lins [1994]
- Wang book for unitarization Wang [2010]
- some of the ideas date back to Penrose [1971]

Exercises

Exercise 19.1 Jones-Wenzl projectors P_0, P_2, and P_3

For two strands one can construct two Jones-Wenzl projectors P_0 and P_2 as shown in Fig. 19.5 and 19.7.

(a) Show that these 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 exercise 2.2.

(c) The three strand Jones-Wenzl projector must be of the form shown in the figure 19.11.

The coefficients $\alpha, \beta, \gamma, \delta, \epsilon$ 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 Figs. 19.12 and 19.13.

Calculate the coefficients $\alpha, \beta, \gamma, \delta$ in P_3. Calculate the quantum dimension d_3 shown in Fig. 19.15.

(d) Choosing $d = \pm\sqrt{2}$ 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: Try putting P_3 within a some simple diagrams and calculate the results.

(e) For the case of $d = \pm\sqrt{2}$ show that, when evaluated in any diagram, $2 \times 2 \notin 2$. In other words, prove Fig. 19.18.

Exercise 19.2 More General Jones-Wenzl Projectors

(a) A Jones-Wenzl projector for n strands is defined both by $P_n^2 = P_{n+1}$ as well as by being orthogonal to P_0 analogous to Fig. 19.13. Assuming these properties are satisfied for P_n show that they are satisfied for P_{n+1} given by Fig. 19.19. Hint: Use the fact that connecting up a single string from P_{n+1} from top to bottom as in Fig. 19.29 must give something proportional to P_n (Why?).

(b) Using Fig. 19.19 derive Eq. 19.1. Show that the solution to this equation is given by Eqs. 19.2 and 19.3. Confirm the condition for d_n to vanish given in Eq. 19.4.

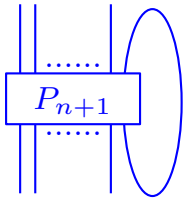


Fig. 19.29 This figure, with n strands going in the bottom, and n strands coming out the top, must be proportional to P_n.

Exercise 19.3 Theta Diagram

(a) Show $\Delta(a+1, a, 1) = d_{a+1}$. Hint: Use Fig. 19.29.

(b) More generally show $\Delta(a+k, a, k) = d_{a+1}$. Hint: Generalize Fig. 19.29 to the case where k strands are connected in a loop from the top to the bottom.

Exercise 19.4 F-matrix diagrammatics

Using the diagrammatic algebra, determine $F_{12\beta}^{21\alpha}$ and $F_{21\beta}^{21\alpha}$ for arbitrary d. Confirm that your results are unitary matrices.

Exercise 19.5 Twists of Kauffman Anyons

Use the Kauffman bracket rules to calculate θ_a for the P_a kauffman anyon. Show that

$$\theta_a^* = (-1)^a A^{-a(a+2)}$$

Hint: Try $a = 2$ then $a = 3$ to figure out the pattern.

20 Anyons from Groups

In this chapter we will use the structure of discrete groups to build rules for anyon theories[1]. There are two general approaches we will consider. In section 20.1 we will label our diagrams with group elements whereas in section 20.2 we will label diagrams with group representations. As we have done previously we will in each case consider planar diagrammatic algebras first before considering braiding.

[1] In chapter *** (also section ***) we discuss another construction of an anyon theory from a discrete group, known as the quantum double construction. We defer discussion of that construction for now.

20.1 Fusion as Group Multiplication

One way to construct a wide variety of consistent planar diagrammatic algebras is to construct our fusion rule based on the structure of a group. In this approach we consider a discrete group G, and each element $g \in G$ is a particle type with the identity element I of the group being the vacuum.

Fusion rules follows the rules for group multiplication. That is, for $g, h \in G$

$$g \times h = gh$$

which we draw as shown in Fig. 20.1.

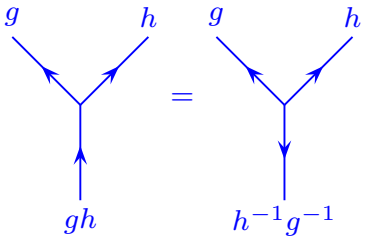


Fig. 20.1 Fusion is defined by group multiplication. On the right we show the three particles orientied as all leaving the vertex. With this orientation when the three particles are multiplied together in clockwise order, they should fuse to the identity $gh(h^{-1}g^{-1}) = h(h^{-1}g^{-1})g = (h^{-1}g^{-1})gh = I$.)

Since $gg^{-1} = g^{-1}g = I$, antiparticles are given by the inverse elements in the group, or $\bar{g} = g^{-1}$. This means that in a diagram we may reverse an arrow if we invert the group element as shown in Fig. 20.2.

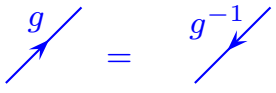


Fig. 20.2 Reversing an arrow inverts the group element.

Let us consider diagrams where each line is labeled by a group element $g \in G$. Reversal of a line corresponds to inversion of the group element as shown in Fig. 20.2 analogous to reversing an arrow in order to turn a particle into its antiparticle.

In cases where the group is abelian so that $g \times h = gh = hg = h \times g$ which is what we required for fusion of particle types in section 8.1 above. In section 20.1.3 we will consider the possibility of using nonabelian groups, but for now we will assume the group is abelian. We thus have fusion rules given by group multiplication

$$N_{g,h}^{a} = \delta_{a,gh} = \delta_{a,hg}$$

Since the result of any fusion is always uniquely defined by group multiplication (one never has a sum on the right hand side, such as $g \times h = a + b$), the quantum dimension of every particle is $\mathsf{d}_g = 1$ meaning the Hilbert space size does not grow with the number of particles.

An example of a planar diagram with this type of group multiplication is shown in Fig. 20.3.

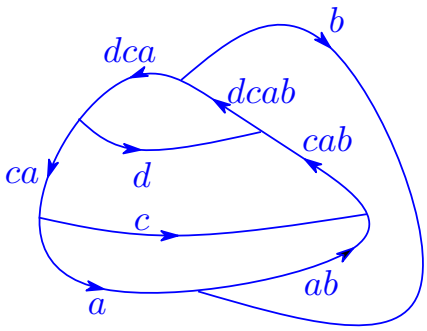


Fig. 20.3 A planar diagram with fusion being defined as group multiplication. For each vertex, if all arrows are pointed out of the vertex, then going around the vertex clockwise, the group elements multiply to the identity, as shown in Fig. 20.1.

20.1.1 Group Cohomology

We now have the task of trying to construct consistent F-matrices for our planar diagram algebra. This is an extremely well studied problem in the field of *group cohomology*.[2]

[2]Group cohomology is a very general framework which we will not delve into more than is necessary. However, it is worth knowing that it enters prominently in a number of topological theories.

Consider a general group G. A so-called 3-cocyle of the group is given by a function of three variables $\omega(a,b,c)$ where $a,b,c \in G$ that satisfies

$$\omega(a,b,c)\omega(a,bc,d)\omega(b,c,d) = \omega(ab,c,d)\omega(a,b,cd) \tag{20.1}$$

Generally we will consider cases of ω being a $U(1)$ valued complex phase. In group cohomology notation we say that

$$\omega \in H^3(G,U(1)) \tag{20.2}$$

Eq. 20.1 may look obscure, but it is actually just a translation of the pentagon equation! Let us make the identification, in the notation of chapter 9,

$$[F^{a,b,c}_{(abc)}]_{(ab),(bc)} = \omega(a,b,c)$$

So that we have diagrammatically

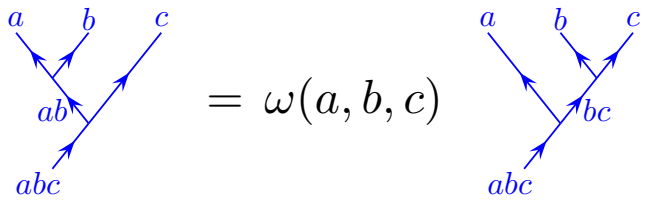


Fig. 20.4 The 3-cocycle is precisely an F-matrix. Compare to Fig. 9.1.

Examining the pentagon equation Eq. 9.7 and Fig. 9.7 we see that this is precisely the same as Eq. 20.1 in a different language. Note that

there is no sum over indices here (like the sum over possible elements h in Eq. 9.7) since the fusion of any two group elements always gives a unique group element as an outcome.

As with F-matrices, it is possible to choose different gauges (See section 9.4). In particular given a 3-cocycle (ie., a solution of the pentagon equation) we can multiply each a, b vertex by a phase $u(a, b)$ as shown in Fig. 20.5 to transform the cocycle by

$$\omega(a,b,c) \to \frac{u(a,bc)u(b,c)}{u(a,b)u(ab,c)}\,\omega(a,b,c). \tag{20.3}$$

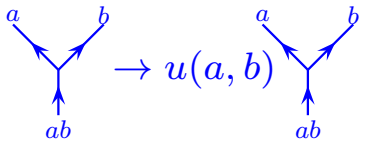


Fig. 20.5 We have the freedom to make a gauge transform of a vertex by multiplying by a phase $u(a, b)$.

By making such a gauge transform we generate additional solutions of the pentagon equation. We view different solutions which are gauge transforms of each other as being physically equivalent, We will typically work with just one representative 3-cocycle for each equivalence class by choosing a convenient gauge. It is useful to always work with a so-called normalized gauge, where $\omega(a, b, c) = 1$ whenever $a = I$ or $b = I$ or $c = I$. (I.e, fusing with the vacuum gives no phase). Further we want to only consider gauge transforms that maintain this normalized gauge, so we must insist on $u(I, g) = u(g, I) = u(I, I) = 1$. Given this restriction to normalized gauge, however, one still has a large additional gauge freedom.

The 3-cocycle (pentagon) equation Eq. 20.1 typically will have more than one gauge-inequivalent solution. Further, if we have two different 3-cocycles ω and ω', we may multiply these together to generate another solution $\omega\omega'$ and we may invert ω to generate another solution. Thus, the space of 3-cocycles $H^3(G, U(1))$ in Eq. 20.2 is itself a group, known as the *third cohomology group of G with coefficients in $U(1)$*.

A trivial 3-cocycle $\omega(a, b, c) = 1$ for all $a, b, c \in G$ is always possible. In this case all diagrams have value 1. However, for any group (beyond the trivial group with only one element), there are always other possible 3-cocycles as well. Such 3-cocycles and group cohomology in general have been studied extensively in the mathematics and physics communities and it is possible to simply look up the form of the possible 3-cocyles. (See the end of the chapter for good references).

While all 3-cocycles provide a solution to the pentagon equation, they do not always allow for full isotopy invariance as discussed in chapter 16. Indeed, for any 3-cocycle ω, we will need to check whether it satisfies all the requirements for full isotopy invariance. For example, if we want to be able to freely turn up and down legs of a vertex as shown in Fig. 20.6.

Thus for full isotopy invariance (and allowing for d both $+1$ and -1) we need to have

$$s(a,b)\omega(a,a^{-1},b) = 1 \tag{20.4}$$
$$s(a,b)\omega(a,b^{-1},b) = 1 \tag{20.5}$$

for all a, b in the group with

$$s(a,b) = \begin{cases} -1 & d_a = d_b = -1 \\ +1 & \text{otherwise} \end{cases} \tag{20.6}$$

While this condition seems quite restrictive, the gauge freedom Eq. 20.3 allows us often to achieve this.

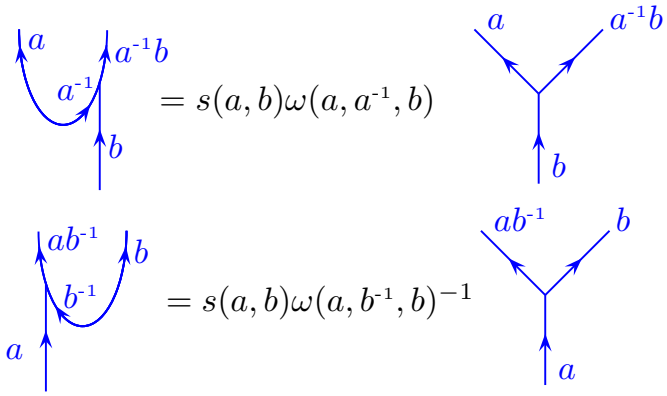


Fig. 20.6 Turning up and down relations (analogous to Fig. 14.10). The prefactor s comes from the proper interpretation of the sign of the $\sqrt{d}$ factors in 14.10. See section 14.2.1.

A further item to note is that the Frobenius-Schur indicator for a particle is

$$d_a = \omega(a, a^{-1}, a)$$

and if a is self-dual ($a = a^{-1}$) this is the gauge invariant Frobenius-Schur indicator. If we can do so, we will try to choose a gauge for our 3-cocyles such that we have full isotopy invariance (this is not always possible) by choosing negative d_a but keeping ϵ_a positive as discussed in section 14.2.1.

20.1.2 Simple Examples with $G = \mathbb{Z}_N$

For example, let us take a simple case of the group $G = \mathbb{Z}_N$, the group of integers modulo N with the group operation being addition modulo N. Since this group is abelian, we have $g \times h = h \times g$ as we want for fusion of particle types as described in chapter 8.

The inequivalent 3-cocyles of the group $\mathbb{Z}_N$ can be written as (See references at the end of the chapter)

$$\omega(a,b,c) = \exp\left(\frac{2\pi i p}{N^2} a(b + c - [b+c]_N)\right) \tag{20.7}$$

where here $a, b, c \in 0, \ldots, N-1$, and the brackets $[b+c]_N$ means $b+c$ modulo N where the result is chosen to lie in the range $0, \ldots, N-1$. Here the index p is an integer in the range $0, \ldots, N-1$ describing the

N different gauge-inequivalent 3-cocyles.

The trivial 3-cocyle is given by $p = 0$ which gives $\omega = 1$ always. The nontrivial 3-cocycles are more interesting.

$\mathbb{Z}_2$

Lets consider the simple case of $\mathbb{Z}_2$ fusion rules. Here the group elements are $g = 0, 1$ and the group operation is addition modulo 2. One has the trivial 3-cocycle $p = 0$ in Eq. 20.7, giving $\omega = 1$, or all F matrix elements equal to 1, which we identify as being exactly the same as the $d = 1$ loop gas from section 18.1.1.

The only nontrvial 3-cocycle is the $p = 1$ case. Here, using Eq. 20.7 we determine the 3-cocycle is of the form

$$\omega(a, b, c) = \begin{cases} -1 & a = b = c = 1 \\ +1 & \text{otherwise} \end{cases} \tag{20.8}$$

We recognize this as being exactly the case of the $d = -1$ loop gas from section 18.1.2 (This translates to saying that the F-matrix is -1 if and only if all four incoming legs a, b, c and abc are in the 1 state as in Eq. 18.10, and note that abc here means multiplication with the group operation so is really $(a + b + c) \bmod 2$.).

$\mathbb{Z}_3$ and beyond

Generalizing the $\mathbb{Z}_2$ fusion to $\mathbb{Z}_3$, we now have $g = 0, 1, 2$ with the group operation being addition modulo three. In this case we have three different 3-cocycles, the trivial 3-cocycle ($p = 0$ in Eq. 20.7) and two nontrivial 3-cocycles ($p = 1$ and $p = 2$ in Eq. 20.7).

While these nontrivial cocycles provide a valid solution to the pentagon equation 20.1 (or Eq. 9.7) they are not in a form where they enjoy full isotopy invariance. One can use gauge transforms Eq. 20.3 to try to put the cocyles in different forms, but it is not possible to find a gauge where both Eq. 20.4 and Eq. 20.5 are satisfied at the same time[3]. Nonetheless they still provide a consistent planar diagrammatic algebra, although not a fully isotopy invariant one. Thus the only isotopy invariant case is the trivial cocycle $p = 0$. We will discuss the possible braidings that are consistent with this planar algebra when we discuss the trivial cocycle below.

One can further show that $\mathbb{Z}_N$ for any odd N is similar to the case of $\mathbb{Z}_3$: the only fully isotopy invariant case is the trivial cocycle $p = 0$. See exercise 20.2. And this theory has N solutions of the hexagon equaation. These are given in detail in appendix 20.4.

For $\mathbb{Z}_N$ with N even, the situation is slightly more interesting. Here, there are two possible values of p which can give an isotopy invariant solution: $p = 0$ (the trivial cocycle similar to what we found in the N odd case) and $p = N/2$, which is analogous to the nontrivial cocycle we found for the $\mathbb{Z}_2$ case. (See exercise 20.2). For each of these values of p there are N solutions of the hexagon equation. All of these cases are

[3] To see that it is not possible to achieve full isotopy invariance note that from Eq. 20.4 and 20.5, isotopy invariance requires $\omega(1, 1, 2) = \pm 1$ and $\omega(1, 2, 2) = \pm 1$. However, for $N = 3$, the product $\omega(1, 1, 2)\omega(1, 2, 2)$ is gauge invariant, and it is only ± 1 for the case of $p = 0$. See exercise 20.2.

discussed in detail in appendix 20.4. Again, for other values of p we have a perfectly consistent planar diagrammatic algebra, although not a fully isotopy invariant one (no matter what gauge we choose).

Other abelian groups

Abelian groups are always of the form $\mathbb{Z}_{N_1} \times \mathbb{Z}_{N_2} \times \ldots$ with some number of factors of cyclic groups $\mathbb{Z}_N$. We can look up the cocycles for such groups in, for example, de Wild Propitius [1995] or Hu et al. [2013] or a book on group cohomology! Note that the variety of different possible cocycles grows when there are multiple $\mathbb{Z}_N$ factors. We will not pursue these theories further here[4].

[4]It looks like a fun exercise to explore this!

20.1.3 Using Nonabelian Groups?

In the case where the group is nonabelian we deviate from what was done when we discussed fusion of particle types in section 8.1 above. In the discussion of fusion of particle types, we have always assumed $g \times h = h \times g$ and with a nonabelian[5] group gh may not be the same as hg.

[5]We have a bit of a language difficulty here. Here we use the word *nonabelian* to mean when $g \times h \neq h \times g$ whereas previously (See section 8.2) we used nonabelian to describe fusion rules where there is more than one fusion channel, such as $g \times h = a + b + \ldots$.

Why did we insist in chapter 8 that particle fusion should satisfy $g \times h = h \times g$? If we think about particles living in three dimensions, when we bring two particles, g and h together, looking at the system from one angle it looks like g is to the right of h but looking at the two particles from another angle, it looks like h is to the right of g. Thus there is no way to decide whether the pair fuses to gh or hg.

However, if we are only concerned with a planar diagram algebra (or a diagram algebra on the surface of sphere) then there is no ambiguity! The surface we are considering is assumed to be oriented so we can always unambiguously decide which particle is clockwise of which other particle at a vertex. Thus we can make the general rule that for a vertex to be an allowed fusion, the three particles *leaving* the vertex must multiply in clockwise order to the identity as shown in the right of Fig. 20.1. Thus, at least for planar diagrams we can generalize our rules for particle fusion to allow non-commutative fusions.

All of the figures we have drawn in this section (Fig. 20.1 – Fig. 20.6) have been drawn so as to be consistent with our rule for nonabelian groups — that is, if all of the arrows are outgoing, when you multiply the group elements clockwise around the vertex you obtain the identity.

For definiteness we discuss the example of the nonabelian group S_3 in the appendix section 20.5.

20.2 Fusion of Group Representations: Rep(G)

Another way to construct a consistent planar diagrammatic algebra is to work with representations of discrete groups[6]. Suppose we have irreducable representations R_i of a group G. A tensor product of two of these irreducible representations will necessarily decompose into a direct

[6]To remind the reader, each discrete group has a finite number of irreducible representations, and any representation of the group can be decomposed into a direct sum of irreducible representations. See section 33.2.4.

sum of irreducible representations. I.e., we have[7]

$$R_a \otimes R_b \simeq R_c \oplus R_d \oplus \dots \tag{20.9}$$

with the sum on the right hand side being finite. We thus propose to label a particle type for our diagrammatic algebra with an irreducible group representation, and have the fusion relations be given by these tensor product decompositions. Thus we interpret the tensor produt equation Eq. 20.9 as a particle fusion relation

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

and accordingly a particle a's corresponding to representation R_a has antiparticle $\bar{a}$ corresponding to the dual representation which we write as $R_{\bar{a}} = R_a^*$. This fusion category (this set of fusion rules with the associated F-matrices) using the representations of the group G is known as Rep(G).

It is fairly easy using some tricks of group theory to determine the fusion rules for discrete group representations. Recall that a representation R is a homomorphism[8] from each group element g to a matrix $\rho^R_{mn}(g)$ (See section 33.2.4). The trace of the representation matrix is known as its character

$$\chi^R(g) = \mathrm{Tr}[\rho^R(g)]$$

One can either work out the characters of a group explicitly or (much more commonly) just look them up on character tables, which can be found in any group theory book or on the web.

Since $\mathrm{Tr}(ab) = \mathrm{Tr}(ba)$ we have $\chi^R(g) = \chi^R(hgh^{-1})$ meaning that the character depends only on the so-called conjugacy class of the group element g.

Characters combine in fairly simple ways under both direct product and direct sum

$$\chi^{R_a \oplus R_b}(g) = \chi^{R_a}(g) + \chi^{R_b}(g) \tag{20.10}$$
$$\chi^{R_a \otimes R_n}(g) = \chi^{R_a}(g)\chi^{R_b}(g) \tag{20.11}$$

Further we have orthonormality relations for irreducible representations:[9]

$$\frac{1}{|G|}\sum_{g\in G}[\chi^{R_a}(g)]^*\ \chi^{R_b}(g) = \delta_{R_a,R_b} \tag{20.12}$$

where the sum is over all elements g of the group G and $|G|$ is the total number of elements in the group. We can thus deduce the tensor product decomposition[10,11]

$$R_a \otimes R_b \simeq \bigoplus_{c\in \text{ irreps}} N_{ab}^c\, R_c \tag{20.13}$$

[7] If we write $M \otimes N = P$ we mean the following. If M_{ab} is a matrix of dimension m and N_{cd} is a matrix of dimension n then P is defined as $P_{(ac),(bd)} = M_{ab}N_{cd}$ and is of dimension nm. If we write $P = N \oplus M$ we mean that P is block diagonal with blocks N and M. Finally note that the relation in Eq. 20.9 is an isomorphism not an equality. One can choose a basis such that the right hand side is block diagonal, however, this is not the natural basis for the left.

[8] Meaning a mapping where the group operation is preserved: $\rho^R(g_1)\rho^R(g_2) = \rho^R(g_1 g_2)$.

[9] This orthonormality is derived trivially from the grand orthogonality theorem, Eq. 33.3. Since the character $\chi(g)$ is a function of the conjugacy class of g only it is sometimes more convenient to replace the sum over all elements with a sum over classes where we then also include a factor of the number of elements in the class. So the left hand side would read instead

$$\sum_{\text{classes } C} \frac{|C|}{|G|}[\chi^{R_a}(g \in C)]^*\ \chi^{R_b}(g \in C)$$

with $|C|$ meaning the number of elements in class C.

[10] The $\bigoplus$ symbol here means a direct sum of all the arguments. The prefactor N_{ab}^c here means the R_c representation occurs N_{ab}^c times in the direct sum.

[11] We have $R_a \otimes R_b \simeq R_b \otimes R_a$ meaning the two tensor products are isomorphic, but they are not equal. The two matrices have their entries in different places. See the definition in note 7 of this chapter above.

	identity 1 element	rotations 2 elements	reflections 3 elements
trivial rep (I)	1	1	1
sign rep (S)	1	1	-1
2d rep (V)	2	-1	0

Table 20.1 Character table for the group S_3. Notice the orthogonality of rows as defined by Eq. 20.12.

where

$$N_{ab}^c = \frac{1}{|G|} \sum_{g \in G} [\chi^{R_c}(g)]^* \ \chi^{R_a}(g) \chi^{R_b}(g) \tag{20.14}$$

or in our fusion product language

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

Note that in the case where the group is abelian, the representations themselves are also an abelian group (meaning $N_{ab}^c = N_{ba}^c \in \{0, 1\}$ only.)

It is not hard to show (See exercise **) that the quantum dimension of a representation R_a is given by

$$\mathsf{d}_a = \chi^{R_a}(e) \tag{20.15}$$

where e is the identity element of the group.

Example: Representations of S_3

As a simple example, let us consider the representations of the group S_3 which can also be thought of as the symmetries of a triangle. To remind the reader[12] this group has 6 elements which can be written in terms of two generators X (a reflection) and R (a rotation) with multiplcation rules $X^2 = R^3 = e$ and $XR = R^{-1}X$ with e the identity. The 6 elements can be written as $e, R, R^{-1}, X, XR, XR^{-1}$. There are three conjugacy classes, which we will call the identity (e), the rotations (R, R^{-1}) , and the reflections (X, XR, XR^{-1}).

There are also three irreducible representations[13]. The group has a character table as given in table 20.1. It is then easy to use Eq. 20.14 to determine the fusion laws for the representations, which are given by

$$I \times I = I, \qquad I \times S = S, \qquad I \times V = V \tag{20.16}$$
$$S \times S = I, \qquad S \times V = V \tag{20.17}$$
$$V \times V = I + S + V \tag{20.18}$$

from which we see that I plays the role of the vacuum particle. Just as an example, let us consider Eq. 20.18. From the character table we have $\chi^V = (2, -1, 0)$ and so $\chi^{V \otimes V} = \chi^V \chi^V = (4, 1, 0) = (1, 1, 1) + (1, 1, -1) +$

[12] The group S_3 is also sometimes known as the dihedral group with 6 elements, often denoted D_3 or sometimes D_6. See section 33.2 for a few more details of this group.

[13] The number of irreducible reps is always equal to the number of conjugacy classes.

class	1	-1	$\{\pm i\sigma_x\}$	$\{\pm i\sigma_y\}$	$\{\pm i\sigma_z\}$
elements	1	1	2	2	2
I	1	1	1	1	1
R_x	1	1	1	-1	-1
R_y	1	1	-1	1	-1
R_z	1	1	-1	-1	1
S	2	-2	0	0	0

Table 20.2 Character table for the group $\mathbb{Q}_8$. Notice the orthogonality of rows as defined by Eq. 20.12.

$$(2,-1,0) = \chi^I + \chi^S + \chi^V.$$

Example: Quaternion Group $\mathbb{Q}_8$

The quaternion group[14] can be defined as the eight two-by-two matrices $\pm\mathbf{1}, \pm i\sigma_x, \pm i\sigma_y, \pm i\sigma_z$. The group has five conjugacy classes 1, -1, $\{\pm i\sigma_x\}$, $\{\pm i\sigma_y\}$, $\{\pm i\sigma_z\}$, and correspodingly five representations. The character table is given in table 20.2. From the character table it is easy to use Eq. 20.14 to derive the nontrivial fusion rules (Again I plays the role of the vacuum particle and we do not write its fusions)

$$R_i \times R_i = I \qquad i = x, y, z \tag{20.19}$$
$$S \times R_i = S \qquad i = x, y, z \tag{20.20}$$
$$R_x \times R_y = R_z \qquad \text{(and cyclic permutations)} \tag{20.21}$$
$$S \times S = I + R_x + R_y + R_z \tag{20.22}$$

Note that S is the 2-dimensional representation given by the defining two-by-two matrices.

[14] The quaternions were famously discovered by Hamilton. He was so excited by this discovery that he carved them into the stone of Brougham (Broom) Bridge in Dublin. There is a plaque there today to commemorate this event.

20.2.1 F-Matrices

With a bit of work, the F-matrices (often known as $6j$ symbols in this context) can also be derived using group theoretic methods. In general this can be a bit complicated but the principle is straightforward group theory. As usual we should think of $F^{bad}_{ec_f}$ as a basis transform (See Fig. 16.3). In this case it is convenient to think of the process of b, a, e and c fusing to the identity in different ways or equivalently, the tensor product of R_b, R_a, R_e and R_c fusing to the identity representation.

(1) Consider $R_a \otimes R_b \simeq \bigoplus_{\bar{d}} N^{\bar{d}}_{ab} R_{\bar{d}}$ and fuse with $R_c \otimes R_e \simeq \bigoplus_d N^d_{ce} R_d$. The resulting representations, R_d and $R_{\bar{d}}$ then fuse together to form the identity representation. Such a process corresponds to the diagram in the top of Fig 20.7 (same as the left of Fig. 16.3).

(2) Consider instead $R_b \otimes R_c \simeq \bigoplus_{\bar{f}} N^{\bar{f}}_{bc} R_{\bar{f}}$ and fuse with $R_a \otimes R_e \simeq \bigoplus_f N^f_{ae} R_f$, and finally fuse $R_{\bar{f}}$ and R_f to form the identity representation. Such a process corresponds to the diagram on bottom of Fig 20.7 (same as the right of Fig. 16.3).

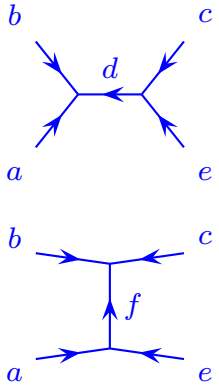


Fig. 20.7 Fusion of representations. Compare to the diagrams in Fig. 16.3. The top figure can be thought of as the d component of $R_a \otimes R_b \simeq \bigoplus_{\bar{d}} N_{ab}^{\bar{d}} R_{\bar{d}}$ fused with $R_c \otimes R_e \simeq \bigoplus_d N_{ce}^d R_d$ to form the identity. The bottom figure can be thought of as the f component of $R_b \otimes R_c \simeq \bigoplus_{\bar{f}} N_{bc}^{\bar{f}} R_{\bar{f}}$ fused with $R_a \otimes R_e \simeq \bigoplus_f N_{ae}^f R_f$ to form the identity.

Both of these processes correspond to fusion of the four representations to the identity. The first, we might say is the identity component of $(R_b \otimes R_a) \otimes (R_e \otimes R_c)$ whereas the second is the identity component of $(R_b \otimes R_c) \otimes (R_a \otimes R_e)$. While these two tensor products are isomorphic, they are expressed in a different basis (see note 11 above). To find the F-matrix $F_{ec_f}^{bad}$ matrix relating these bases we simply have to find the overlap between the d contribution to the overall identity representation in case (1) above with the f contribution to the overall identity representation in case (2).

20.2.2 Some Simple Braidings for Rep(G)

So far we have only discussed consistent fusion of representations – i.e, fusion rules that will satisfy the pentagon equation. Given a set of F-matrices, we can then look for braidings, or R-matrices that satisfy the hexagon. We already know that there will typically be multiple solutions[15] of the hexagon. Some of the braidings, however, can be stated easily for any group G, as we now discuss.

Trivial Braidings: Bosons

It is always the case that the trivial braiding

$$R_{a\times b}^{a,b} = 1 \tag{20.23}$$

provides a solution[16] of the hexagon for Rep(G) for any group G. If we choose this braiding we are describing particles that are bosons — i.e., having trivial braiding[17] and trivial spin $\theta_a = 1$ — with internal quantum numbers given by the representations of the group G.

Fermions and Bosons

If it so happens that G contains an central element[18] z such that $z^2 = e$ with e the identity of the group, then each representation R_a can be assigned a degree $p(a)$ which is 1 or 0 depending on whether z acts as the identity in the representation or acts as -1 in the representation[19] A consistent braiding is then given by (See exercise ***)

$$R_{a\times b}^{a,b} = (-1)^{p(a)p(b)} \tag{20.24}$$

In other words, we have declared a particle a to be if bosonic if $p(a) = 0$ (if $\rho_{R_a}(z)$ acts as 1) or fermionic $p(a) = 1$ (if $\rho_{R_a}(z)$ acts as -1).[20]

[15] Recall that if G is abelian, then the representations of G are isomorphic to the group G itself. In which case Rep(G) is just the trivial cocylce of G. As we discussed in section 20.1.2 and 20.4 the trivial cocycle of $\mathbb{Z}_N$ has N different braidings.

[16] Not every set of fusion rules is consistent with the trivial braiding (For example, the Ising fusion rules, as discussed in section 18.3.1, are not consistent with a trivial braiding). The fact that Rep(G) always allows a trivial braiding is a subtle, but straightforward statement about using tensor products of representations to represent fusion of particle types. We are in essence representing the braid operator by $B(R_a \otimes R_b) = R_b \otimes R_a$, i.e., we just re-order the tensor factors without adding any phase. Realizing that the F-matrices are also just relating different re-orderings of tensor factors, this tells us that the two paths through the hexagon must give the same result.

[17] Note that R_c^{ab} is gauge dependent, so Eqs. 20.23 and 20.24 imply a gauge choice. However, $R_c^{ab}R_c^{ba}$ is gauge independent, and θ_a is gauge independent.

[18] "Central" means that z commutes with all the elements of the group. I.e., $zg = gz$ for all $g \in G$.

[19] To see this consider applying z in the representation twice $\rho^R(z)\rho^R(z) = \rho^R(z^2) = \rho^R(e)$. This means that $\rho^R(z)$ can only act as the identity or as minus the identity.

[20] Sometimes this braided solution to the hexagon is called Rep(G, z) compared to the fully bosonic solution which can then be called Rep(G, e). I.e., if you set $z = e$, the identity, then you set all particles to be bosons.

The fermionc particles have $\theta_a = -1$ whereas the bosonic particles have $\theta_a = +1$.

An nontrivial example of $\text{Rep}(G, z)$ is given by the quaternion group $\mathbb{Q}_8$ where we choose $z = -1$. It is easy to check that this makes the R_i representations bosons, but the S representation is a fermion (note the negative sign on the character $\chi^S(-1)$).

20.2.3 Continuous (Lie) Group Representations?

One can imagine that instead of looking at the representations of discrete groups, one considers instead the representations of Lie groups (See section 33.2.3). For example, the different representations of the group $SU(2)$ are the different values of the spin quantum number j, and these fuse together with the usual angular momentum addition rules. Further, the F-matrices are (up to a normalization) precisely what we call $6j$ symbols of angular momentum addition.

While such a scheme makes a perfectly good planar diagrammatic algebra, the problem is that there are an infinite number of different representations (For the case of $SU(2)$ for example, the angulular mommentum j can inifinitely large) and this violates our rule of having a finite number of "particle types" for our diagrammatic algebra. Such algebras can be problematic when used for physical purposes (For example, as we will see in section 21.3 using a diagrammatic algebra with an infinite number of representations for construction of a TQFT results in divergences). Schemes have been constructed to regularize such a diagrammatic algebra and arrange that only a finite number of representations ever occur – which are often known as "deformations" of the Lie algebra representation[21]. The most common such deformations correspond precisely to the particle types of a corresponding Chern-Simons theory at some finite level. For example, in the case of $SU(2)$, one can consider $SU(2)_k$ Chern-Simons theory which has deformed F-matrices such that angular momentum $j = 0, 1/2, \ldots k/2$ can occur, but one never gets any higher angular momenta.

[21] The term "quantum group" is often used. Be warned that a quantum group is not a group.

20.3 Parastatistics Revisited

Way back in section 3.5.1 we asked why we could not have exotic statistics in 3+1 dimensions. While there are nontrivial representations of the permutation group that would satisfy the quantum mechanical composition rule, we stated that additional constraints — such as particle creation and annihilation and locality — limits us to just bosons and fermions. We are now at the point where we can discuss exactly what we mean by this.

The structure we have built up for anyons in 2+1 dimensions is that of a braided unitary category: a set of particles with fusions, F-matrices satisfying the pentagon, and R-matrices satisfying the hexagon. If we try to do something similar in 3+1 dimensions we will no longer can have nontrivial braiding of world lines since, as discussed in section 3.3.2, in

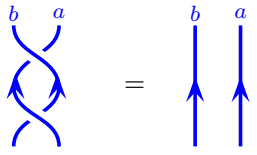


Fig. 20.8 In 3+1 dimensions these two pictures are topologically equivalent. Thus all particles are transparent. This implies Eq. 20.25.

3+1 dimension no knots can be formed in one-dimensional world lines. Thus, we must impose the restriction shown in Fig. 20.8 that all particles are transparent. In equations this can be stated as

$$R_c^{ab} R_c^{ba} = 1 \tag{20.25}$$

for all a, b, c such that $N_{ab}^c > 0$. If the condition Eq. 20.25 holds and yet we have a solution of the hexagon equation, we say we have a *symmetric tensor category*[22]. Thus if we are to construct an anyon theory with point particles in 3+1 dimensions, it must be described by a symmetric tensor category.

In fact we have already given two examples of symmetric tensor categories, in section 20.2.2: (a) the theory Rep(G) (or Rep(G, e)) which describes bosons having internal quantum numbers given by the represntations of the group G, and (b) the theory Rep(G, z) where some of the particles are instead declared fermions depending on how their correspodning representation transforms under the action of the element z. The crucial theorem we mentioned in section 3.5.1, originally due to Doplicher and Roberts (See also Müger [2007]; Deligne [2002]), is that there are no other possibilities: Any symmetric tensor category is equivalent to Rep(G) =Rep(G, e) if it has no fermions, or Rep(G, z) if it has fermions. In other words for point particles in 3+1 dimensions, there are only bosons and fermions; nothing else!

[22] A symmetric tensor category is in some sense the exact opposite of a modular tensor category. For a modular tensor category no particles are transparent except the identity, whereas for a symmetric tensor category all particles are transpartent.

20.4 Appendix: Isotopy Invariant Planar Algebras and Anyon Theories from $G = \mathbb{Z}_N$ Cohomology

Here we start with the cocycles for $\mathbb{Z}_N$ given in Eq. 20.7 and we look for isotopy invariant cases. We have two general types of isotopy invariant solutions the $p = 0$ (trivial cocycle) solution, which exists for any N and the $p = N/2$ case which exists for even N only. Let us discuss each of these in a bit more detail.

20.4.1 Trivial Cocycle: $\mathbb{Z}_N^{(n)}$ Anyons

For any $\mathbb{Z}_N$ one can always choose $p = 0$ in Eq. 20.7 which gives the trivial cocycle $\omega(a, b, c) = 1$ for all a, b, c and we correspondingly have $d_a = 1$ for all a.

Recall that these cocycles are really F-matrices, which have now all been set to unity. We now want to determine the possible braidings for this theory by using the hexagon Eq. 13.2. Plugging in $F = 1$ into the hexagon, we obtain[23]

$$R^{c,a}_{[c+a]_N} R^{c,b}_{[c+b]_N} = R^{c,[a+b]_N}_{[c+a+b]_N} \tag{20.26}$$

There are exactly N solutions[24] of this system which we label $n =$

[23] Recall the notation $[a]_N$ means a modulo N.

[24] Examining $\rho_c(a) = R^{c,a}_{[a+c]_N}$ we see that $\rho_c(a)$ is a group representation of the group $\mathbb{Z}_N$, which can only be of the form $\exp(2\pi i p a/N)$ for some p. Finally we invoke the symmetry 13.1.

$0, \ldots, N-1$ which are given by

$$R^{a,b}_{[a+b]_N} = \exp\left[\frac{2\pi i n}{N} ab\right]$$

where ab on the right is actual multiplication, not the group operation which is addition modulo N. The twist factors are (using Eq. 15.2)

$$\theta_a = e^{2\pi i n a^2/N}$$

and the corresponding S matrix is (perhaps easiest derived with Eq. 17.16)

$$S_{a,b} = \frac{1}{\sqrt{N}} \exp\left[\frac{4\pi i n}{N} ab\right]$$

This is a modular theory only for N odd with n and N mutually prime[25]. These theories are sometimes known as $\mathbb{Z}_N^{(n)}$ anyons (See Bonderson [2007]).

[25] For $n = (N-1)/2$ the modular theory matches the Chern-Simons theory $SU(N)_1$ with N odd.

20.4.2 Nontrivial Cocycle: $\mathbb{Z}^{(n)}_{N=2p}$

Here we consider $N = 2p$. The cocycle in Eq. 20.7 again has N consistent solutions of the hexagon equations, given by $n = 0, \ldots, (N-1)$ in the equation (See Bonderson [2007])

$$R^{a,b}_{[a+b]_N} = \exp\left[\frac{2\pi i(n+\frac{1}{2})}{N} [a]_N [b]_N\right]$$

with all $d_a = +1$, resulting in (from Eq. 15.1)

$$\theta_a = \exp\left[\frac{2\pi i(n+\frac{1}{2})}{N} [a]_N^2\right] \tag{20.27}$$

$$S_{a,b} = \frac{1}{\sqrt{N}} \exp\left[\frac{4\pi i(n+\frac{1}{2})}{N} [a]_N [b]_N\right] \tag{20.28}$$

These are known as $\mathbb{Z}_N^{(n+\frac{1}{2})}$ anyon theories for obvious reasons. They are modular when $2n+1$ and p are coprime[26].

These anyon theories, in the gauge given by Eq. 20.7 are not generally isotopy invariant. We can generally make transformations to put these results in potentially simpler forms.

[26] When $n = (N/2 - 1)$ the modular theory matches $SU(N)_1$ with N even, and when $n = 0$ the modular theory matches $U(1)_{N/2}$. Be cautioned that there is some disagreement in the literature as to how you label the level of a $U(1)$ Chern-Simons theory.

Case I: p-odd

With $N = 2p$ and p odd, the p^{th} particle is self dual particle, and has Frobenius-Schur indicator -1, which is a gauge invariant quantity. We will thus need to push this sign onto d in order to have a fully isotopy invariant theory. As discussed in section 14.2.1 we choose a gauge where

$$d_a = (-1)^a$$

(and as usual we are working with all $\epsilon_a = +1$ for an isotopy invariant theory). Note that the composition rule Eq. 14.4 is satisfied. In this gauge the cocycle can be written as[27]

$$\omega(a,b,c) = \begin{cases} -1 & a,b,c \text{ all odd} \\ +1 & \text{otherwise} \end{cases} \tag{20.29}$$

The hexagon equation now takes the form

$$R^{c,a}_{[c+a]_N} R^{c,b}_{[c+b]_N} = \omega(a,b,c) R^{c,[a+b]_N}_{[c+a+b]_N}$$

which compared to Eq. 20.26 introduces a minus sign if all a, b, c, are odd. This system of equations again has N solutions which we index as $\bar{n} = 0, \ldots, N-1$,

$$R^{a,b}_{[a+b]_N} = \exp\left[\frac{2\pi i \bar{n}}{N}[a]_N[b]_N\right](i)^{r(a,b)}$$

where $r(a,b) = 1$ if both a and b are odd and equals zero otherwise. which gives us (using Eq. 15.1 or Eq. 15.2 with $d_a = (-1)^a$ and $\epsilon_a = +1$)

$$\theta_a = e^{2\pi i \bar{n} a^2/N}\ (i)^{t(a)}$$

where $t(a) = 1$ if a is odd and is zero otherwise. It is a short exercise to show that this recovers the correct θ_a, Eq. 20.27, and S-matrix Eq. 20.28 with the mapping $\bar{n} = n - (p-1)/2$. Again, the advantage of using this gauge for the description of $\mathbb{Z}_N^{(n+\frac{1}{2})}$ anyons is full isotopy invariance (with $N = 2p$ and p odd).

[27] If we try to use the same rule for $N = 2p$, with p even where we still set $d_a = (-1)^a$, this actually gives us the trivial cocycle discussed above in a less convenient gauge. Note that the self-dual particle has $d = +1$ in the p even case indicating that we do not need to push any signs onto d.

Case II: p-even

In this case the Frobenius-Schur indicator of the self-dual particle is $+1$ so we can choose to work in a gauge where all $d_a = +1$. While it is possible to make a gauge transform that puts the theory into an isotopy invariant form, the transform is not particularly transparent. For this reason it is often convenient to stay with the gauge given in Eq. 20.7. However, it is not too hard to transform to an isotopy invariant gauge if we would like.

An example of this, let us consider the nontrivial cocycle for the group $\mathbb{Z}_4$. Here we can make a gauge transform (See exercise 20.2) such that

$$\omega(a,b,c) = \begin{cases} -1 & (a,b,c) = (1,1,1);(1,2,3);(2,1,2); \\ & \quad (2,3,2);(3,2,1);(3,3,3) \\ 1 & \text{otherwise} \end{cases} \tag{20.30}$$

and with $d = 1$ for all particles 0,1,2,3. This gives a fully isotopy invariant theory. There are correspondingly 4 solutions of the hexagon given by[28] $n = 0, 1, 2, 3$ with

[28] This n is the same as that of Eq. 20.28

$$R^{a,b}_{[a+b]_4} = \exp\left[\frac{2\pi i(n+\frac{1}{2})}{4}[a]_4[b]_4\right](-1)^{r(a,b)}$$

where $r(a,b) = 1$ for $(a,b) = (1,2),(1,3),(2,1),(3,1)$ only and is zero otherwise.

20.5 Appendix: Cocyles for S_3

To give an example of a non-abelian group, let us look at the case of the group S_3. To remind the reader this group has 6 elements which can be written in terms of two generators X and R with multiplcation rules $X^2 = R^3 = e$ and $XR = R^{-1}X$ with e the identity. The 6 elements can be written as $e, R, R^{-1}, X, XR, XR^{-1}$. Let us write them as $(A,a) = X^A R^a$ with $A = 0,1$ and $a = -1,0,1$. There are 6 independent 3-cocycles described by $p = 0,\ldots 5$ in the equation (See references at the end of the chapter)

$$\omega((A,a),(B,b),(C,c)) = \qquad (20.31)$$
$$\exp\{i\pi pABC\}\exp\left\{\frac{2\pi ip}{9}(-)^{B+C}a\left\{(-)^C b + c - [(-)^C b + c]_3\right\}\right\}$$

where the bracket $[]_3$ indicates modulo 3 where the result is assumed to be in the range $-1,0,1$.

Note that within S_3 there is a $\mathbb{Z}_2$ subgroup consisting of e and X, or $a = 0$ with $A = 0,1$. The first term on the right hands side, $\exp(i\pi pABC)$, matches the two possible 3-cocyles from the $\mathbb{Z}_2$ group. For even p it is the trivial cocycle, whereas for odd p we have a ω being -1 only when A,B,C are all in the 1 state, equivalent to Eq. 20.8. The second factor looks similar to the $\mathbb{Z}_3$ cocycles but only when $C = 0$. Setting $C = 0$ for a moment, the same argument as in the $\mathbb{Z}_3$ case shows that we cannot have full isotopy invariance unless $p = 0$ or $p = 3$, in which case the second factor on the right hand side of Eq. 20.31 is trivial. Thus this case of $p = 3$ gives an isotopy invariant cocyle which essentially ignores the a variable of (A,a) and is equivalent to Eq. 20.8 for the A variables with $d_{(A,a)} = (-1)^A$.

20.6 Details of Working out an F-matrix

Here we give a few more details of how we work out the F-matrix for a group. Other, potentially more systematic schemes, can be found in Refs. Hamermesh [1989]; Buerschaper and Aguado [2009]; Wang et al. [2020] for example.

To impliment our procedure, we work with explicit D dimensonal matrices ρ^R_{ij} for each D dimensional unitary representation R. We extract the identity component of the fusion of the four particles by writing[29]

[29] This is a result of the grand orthogonality theorem Eq. 33.3. If we sum over all group elements we extract only the identity representation.

$$\sum_g \rho^{R_a}(g) \otimes \rho^{R_b}(g) \otimes \rho^{R_c}(g) \otimes \rho^{R_e}(g) \quad = \quad C \sum_d \mathbf{w}_d \mathbf{w}_d^\dagger \tag{20.32}$$

where $\mathbf{w}_d$ is a unit length orthogonal $D_a D_b D_c D_e$ dimensional vector representing the process where a and b fuse to $\bar{d}$ and also c and e fuse to d, and C is an unimportant normalization constant. (Note there is a gauge choice here in choosing the phase of $\mathbf{w}_d$). Similarly we can write

$$\begin{aligned} \sum_g \rho^{R_a}(g) \otimes \rho^{R_c}(g) \otimes \rho^{R_b}(g) \otimes \rho^{R_e}(g) \quad &= \quad C \sum_f \mathbf{z}_f \mathbf{z}_f^\dagger \qquad (20.33) \\ &= \quad C \sum_d P \mathbf{w}_d \mathbf{w}_d^\dagger P^T \end{aligned}$$

where $\mathbf{z}_d$ is a $D_a D_b D_c D_e$ dimensional vector representing the process where a and c fuse to $\bar{f}$ and also b and e fuse to f. In the second line of Eq. 20.33, the matrix P is simply a permutation matrix since the two tensor products are isomorphic and simply have rows and columns appropriately permuted.

The F matrix is then just given by the overlap

$$F^{bad}_{cef} = \mathbf{w}_d^\dagger \cdot \mathbf{z}_f \tag{20.34}$$

The challenge is then simply to extract the correct vectors $\mathbf{w}_d$ and similarly $\mathbf{z}_f$ in Eq. 20.32 and 20.33.

We thus only need to buid up the tensor product in Eq. 20.32 and 20.33 step by step. We have

$$\rho^{R_a}(g) \otimes \rho^{R_b}(g) = \sum_{\bar{d} \in a \times b} \sum_{\alpha,\beta=1}^{D_{\bar{d}}} \mathbf{x}^{\bar{d}}_\alpha \left[\rho^{R_{\bar{d}}}(g)\right]_{\alpha\beta} [\mathbf{x}^{\bar{d}}_\beta]^\dagger \tag{20.35}$$

where the $\mathbf{x}$'s are a set of orthonormal $D_a D_b$ dimensional vectors (both sides of this equation are $D_a D_b$ dimensional matrices)[30]. The particular form of the $\mathbf{x}$ vectors can be extracted using the grand orthogonality theorem Eq. 33.3. Performing the same decomposition for

$$\rho^{R_c}(g) \otimes \rho^{R_e}(g) = \sum_{d \in c \times e} \sum_{\gamma,\delta=1}^{D_d} \mathbf{y}^d_\gamma \left[\rho^{R_d}(g)\right]_{\gamma\delta} [\mathbf{y}^d_\delta]^\dagger \tag{20.36}$$

To find the identity element of the fusion between the tensors in Eq. 20.35 and 20.36 we simply match up the d representations with the $\bar{d}$ representations. Thus we have

$$\begin{aligned} \mathbf{w}_d \mathbf{w}_d^\dagger \quad &\sim \quad \sum_g \sum_{\alpha,\beta,\gamma,\delta=1}^{D_d} \left(\mathbf{x}^{\bar{d}}_\alpha \left[\rho^{R_{\bar{d}}}(g)\right]_{\alpha\beta} [\mathbf{x}^{\bar{d}}_\beta]^\dagger\right) \otimes \left(\mathbf{y}^d_\gamma \left[\rho^{R_d}(g)\right]_{\gamma\delta} [\mathbf{y}^d_\delta]^\dagger\right) \\ &\sim \quad \sum_{\alpha,\beta=1}^{D_d} \left(\mathbf{x}^{\bar{d}}_\alpha [\mathbf{x}^{\bar{d}}_\beta]^\dagger\right) \otimes \left(\mathbf{y}^d_\alpha [\mathbf{y}^d_\beta]^\dagger\right) \qquad (20.37) \end{aligned}$$

[30] In cases where $N^{\bar{d}}_{ab} > 1$ we must take extra care to separate to add the multiple instances of each representation. For simplicity let us assume $N^d_{ab} = 0$ or 1 only.

where in going to the second line we have used the grand orthogonality theorem Eq. 33.3. Both sides of this equation are $D_a D_b D_c D_e$ dimensional matrices with a single nonzero eigenvalue, so it is then trivial to extract $\mathbf{w}_d$. Then extracting $\mathbf{z}_f$ by permuting rows of the vector $\mathbf{w}_d$ as noted in Eq. 20.33, we then determine the F-matrix using Eq. 20.34.

The procedure outlined here is fairly straightforward, although tedious. Equivalent schemes are outlined in Refs. Buerschaper and Aguado [2009] and Wang et al. [2020]. In exercise 20.6 we walk through calculation of F-matrices for the representations of the group S_3 whose fusion rules we worked out in Eqs. 20.16–20.18 above.

As a final comment we note that there is a beautiful expression to calculate the Frobenius-Schur indicator κ for a group representation R

$$\kappa_R = \frac{1}{|G|} \sum_g \chi^R(g^2) \tag{20.38}$$

which gives ± 1 if R is self-dual, and gives 0 for a non-self-dual representation. Recall that the Frobenius-Schur indicator is actually one of the elements of the F matrix (See Eq. 14.3). Derivation of this formula is a standard result in group theory (See, for example, Hamermesh [1989]).

Further Reading

Exercises

Exercise 20.1 Cocycle Equation

(a) Show that the 3-cocycle given by Eq. 20.7 satisfies cocycle condition Eq. 20.1 and thus represents a valid cocycle.

(b) Show that Eq. 20.31 also satisfies Eq. 20.1.

Exercise 20.2 Isotopy Invariance of Cocycles

(a) Show that the cocycle Eq. 20.1 can only represent an isotopy invariant diagram algebra only in the following cases:

- For n an odd integer, only when $p = 0$.
- For n an even integer, only when $p = 0$ or $p = n/2$

(b) For the case of $n = 4$ and $p = 2$, find the gauge transformation that transforms Eq. 20.1 into Eq. 20.30.

Exercise 20.3 Quantum Dimension of a Representation

Prove Eq. 20.15. Hint: Remember that that quantum dimension d_a tells you how the Hilbert space dimension grows as you fuse together the particle a many times. Try fusing together many representations $R^a \otimes R^a \otimes R^a \ldots$ and imagine decomposing the result into irreducible representations using the orthogonality theorem for characters. Note that for characters $\chi(e) \geq \chi(g)$ for e the identity represnentation.

Exercise 20.4 Frobenius-Schur Indicators in Rep(G)

(a) Use Eq. 20.38 to calculate the Frobnius-Schur indicators of the representations for the groups S_3 and $\mathbb{Q}_8$.

(b) The dihedral group with 8 elements, D_8 (sometimes called D_4) is the group of symmetries of a square. Look up the properties of this group. It turns out that it has exactly the same character table as $\mathbb{Q}_8$ (!!). Show that the Frobenius-Schur indicators do not match that of $\mathbb{Q}_8$.

Exercise 20.5 Bosons and Fermions in Rep(G, z)

Let z be a central element of the group G (i.e, $zg = gz$ for all $g \in G$) such that $z^2 = e$, the identity. As in section 20.2.2 for a representation R_a set $p(a) = 0$ if z acts as the idenity in representation R_a and set $p(a) = 1$ if z acts as -1 in representation R_a.

(a) Show that if $R_a \otimes R_b = R_c \oplus \ldots$, then $p(a)p(b) = p(c)$. Hint, conisder the characters $\chi^R(e)$ and $\chi^R(z)$.

(b) Given that setting all particles to bosons (i.e., Eq. 20.23) solves the hexagaon equation, show that Eq. 20.24 also provides a solution to the hexagon equation.

Exercise 20.6 Some F matrix elements for representatons of S_3 [Hard]

Let us consider the simplest nonabelian group S_3,which we discuss in sections 20.5, 11, and 33.2.1.

We remind the reader taht this group has 6 elements which can be written in terms of two generators X and R with multiplcation rules $X^2 = R^3 = e$ and $XR = R^{-1}X$ with e the identity. The 6 elements can be written as e, R, R^2, X, XR, XR^2 which are grouped into conjugacy classes $\{e\}, \{R, R^2\}, \{X, XR, XR^2\}$ (See Table 20.1).

The three representations are as follows: The trivial represenatation has $\rho^I(g) = 1$ for all g in the group. The sign rep has $\rho^S(g) = 1$ for $g \in \{e, R, R^2\}$ and $\rho^S(g) = -1$ for $g \in \{X, XR, XR^2\}$. (Note that since both these reps are one dimensional, they are completely defined by the character table). We write the two dimensional represenation in a unitary form as

$$\rho^V(X) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \rho^V(R) = \frac{-1}{2}\begin{pmatrix} 1 & \sqrt{3} \\ -\sqrt{3} & 1 \end{pmatrix}$$

with $\rho^V(e)$ the identity matrix and all other matrices $\rho^V(g)$ for the other elements g in the group can be generated by using the group multiplication properties. (Do this first, you will need it later!)[31]

Note that we already know the fusion rules for these representations as they are given in Eqs. 20.16–20.18.

In this exercise we will calculate some F-matrix elements by focusing on the most interesting case, where all four incoming lines in Fig. 16.3 are in the two dimensional V representation. Thus we are interested in the unitary matrix F^{VVd}_{VVf}.

(a) Using the grand orthogonality theorem (Eq. 33.3) find the decomposition

$$\rho^V(g) \otimes \rho^V(g) = \mathbf{x}^I[\mathbf{x}^I]^\dagger + \rho^S(g)\mathbf{x}^S[\mathbf{x}^S]^\dagger + \sum_{\alpha,\beta=1}^{2} \mathbf{x}^{\bar{V}}_\alpha \, [\rho^{R_{\bar{d}}}(g)]_{\alpha\beta} \, [\mathbf{x}^{\bar{V}}_\beta]^\dagger$$

Hint: The one dimensional representations are easy to obtain since they can be obtained by

$$(1/|G|) \sum_g \rho^R(g)^* [\rho^V(g) \otimes \rho^V(g)]$$

[31] It may be useful to use a computer to multiply matrices (Mathematica, matlab, octave, and python are all fairly convenient), since there are a lot of matrix manipulations in this problem and a single error will destroy the result.

what remains is the two dimensonal representation.

(b) Given two dimensional matrices A, B, C, D find the permutation matrix such that

$$P(A \otimes B \otimes C \otimes D)P^T = A \otimes C \otimes B \otimes D$$

(c) Use Eq. 20.34 and 20.37 to show that the F matrix is given by

$$F_{VVf}^{VVd} = \frac{1}{2}\begin{pmatrix} 1 & 1 & \sqrt{2} \\ 1 & 1 & -\sqrt{2} \\ \sqrt{2} & -\sqrt{2} & 0 \end{pmatrix}$$

State Sum TQFTs

21

Having learned about planar diagrammatic algebras we are now in a position to explicitly construct a real 3D TQFT[1]. There are several steps in this idea. We start by considering a closed 3D manifold $\mathcal{M}$ which we discretize into tetrahedra (a so called *simplicial* decomposition of the manifold). Next we construct a model, similar in spirit to statistical mechanics, which sums a certain weight over all quantum numbers on all edges of all tetrahedra. The weights being summed are defined in terms of our planar diagrammatic algebra as we will see below. The result of this sum is the desired TQFT partition function $Z(\mathcal{M})$ which we discussed extensively above, and particularly in chapter 7.

This discretization of a manifold into tetrahedra is very commonly used in certain approaches to quantum gravity, which we will discuss in section 21.3.

[1] The input for the construction in chapter will be a planar diagram algebra — *we do not have to specify any sort of R-matrix or braiding.* It is a bit surprising that one only needs a planar algebra to make a 3D TQFT! In section 21.2 we will input a spherical tensor category whereas in section 21.4 we will input a group and a 3-cocycle.

21.1 Simplicial Decomposition and Pachner Moves

We start by considering a so-called simplicial decomposition of our manifold. Such decompositions can be made of smooth manifolds in any number of dimensions[2].

[2] It is interesting (but beyond the scope of this book) that manifolds exist in dimension $d \geq 4$ that cannot be smoothed, and cannot be decomposed into simplicies.

21.1.1 Two Dimensions

As a warm up let us think about two-dimensional manifolds. In two dimensions, the elementary 2-simplex is a triangle, so this decomposition is the familiar idea of triangulation shown in Fig. 21.1.

Since we are only concerned with the topology of the manifold, not the geometry, the precise position of vertex points we use is irrelevant — only

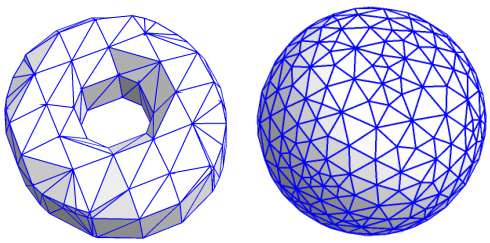


Fig. 21.1 Some triangulations of 2-manifolds

the connectivity of the points is important, i.e, the topological structure of the triangulation network. Furthermore, a particular manifold, like a sphere, can be triangulated in many different ways. It turns out that any two different triangulations can be related to each other by a series of elementary "moves" known as two-dimensional Pachner moves[3,4], which are shown in Figs. 21.2 and 21.3.

[3] I encourage you to play with these two moves and see how you can restructure triangulations by a series of Pachner moves.

[4] It is interesting to note that a Pachner moves can be thought of as viewing a 3D tetrahedron from two opposite directions. We can thus think of 2D Pachner moves as a *cobordism* (See chapter 7) in 3D between a surface triangulated with the initial triangulation and a topologically equivalent surface triangulated with the final triangulation.

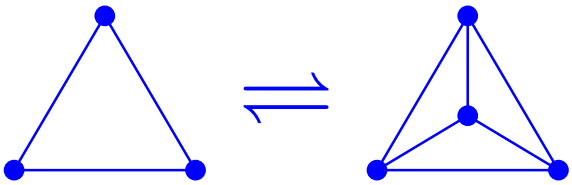


Fig. 21.2 The 1-3 Pachner move in two dimensions corresponds to adding or removing a point vertex from the triangulation. This turns one triangle into three or vice-versa.

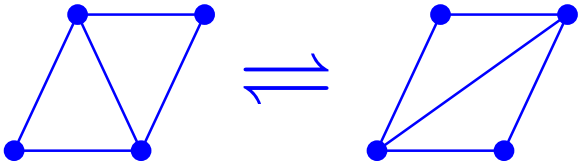


Fig. 21.3 The 2-2 Pachner move in two dimensions corresponds to replacing two adjacent triangles with two complementary triangles. This turns two triangles into two different triangles.

Thus if we want to construct a manifold invariant (like $Z(\mathcal{M})$ we discussed in chapter 7) with a manifold represented in terms of a triangulation we only need to find some function of the triangulation that is invariant under these two Pachner moves.

21.1.2 Three Dimensions

The story is quite similar in three dimensions. Since we have been focused on 2+1 dimensional TQFTs we will mostly discuss three-dimensional manifolds. We discretize any closed three-dimensional manifold[5] by breaking it up into tetrahedra (otherwise known as three-dimensional simplices). Any two discretizations are topologically equivalent to each other if they can be related to each other by a series of three-dimensional Pachner moves[6], which are shown in Figs. 21.4 and 21.5. Again, the key point here is that if we can find some function of the the network structure that is invariant under the Pachner moves, we will have constructed a topological invariant of the manifold.

[5] For now let us focus on closed manifolds. We briefly discuss manifolds with boundary in section 21.2.2.

[6] Analogous to the 2D case (see note 4 above), the 3D Pachner moves can be thought of as viewing a 4D-simplex (a so-called pentachoron) from two opposite directions.

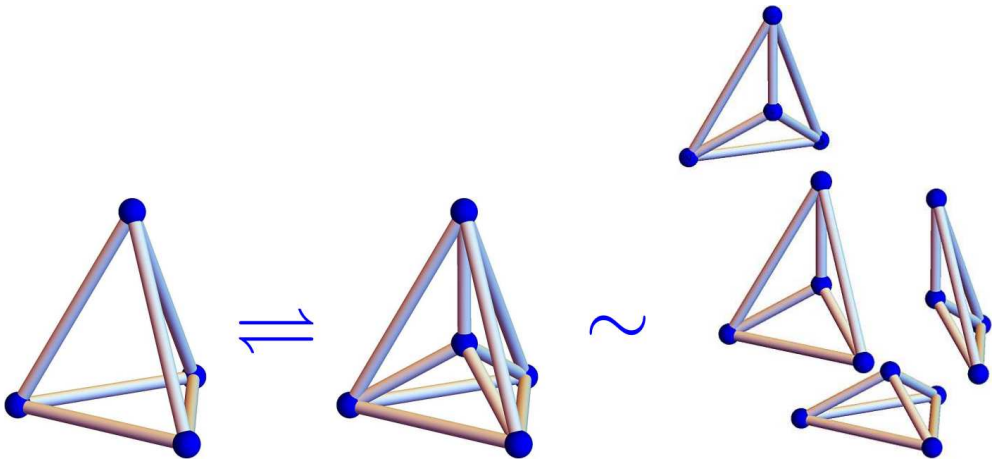


Fig. 21.4 The 1-4 Pachner move in three dimensions corresponds to adding or removing a point vertex to the tetrahedon decomposition. This turns a single tetrahedron into four or vice versa. On the far right we show the four tetrahedron separated for clarity.

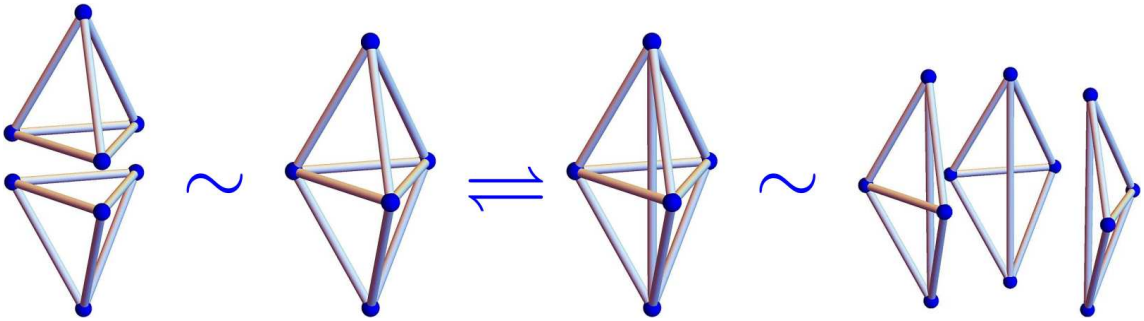


Fig. 21.5 The 2-3 Pachner move in three dimensions corresponds to re-splitting a double tetrahedron (left) into three tetrahedron (right). This turns a single tetrahedron into four or vice versa. On the far left we show the two tetrahedra separated for clarity; and on the far left we have the three tetrahedra separated for clarity.

21.2 The Turaev-Viro State Sum

The idea of the Turaev-Viro state sum is to build a 3D manifold invariant from one of the planar diagrammatic algebras we have been discussing in chapters 8-20.

First, let us choose any particular planar diagrammatic algebra. We take any decomposition of an orientable three dimensional manifold into tetrahedra. Let each edge of this decomposition be labeled with one of the quantum numbers (the particle labels) from the diagrammatic algebra[7]. We then consider the following sum

$$Z_{TV}(\mathcal{M}) = \mathcal{D}^{-2N_v} \sum_{\text{all edge labelings}} W(\text{labeling}) \qquad (21.1)$$

[7] As we have been doing all along, when we label an edge with a quantum number we must put an arrow on the edge unless the particle type is self-dual.

where N_v is the number of vertices in the decomposition, and

$$\mathcal{D} = \sqrt{\sum_n |d_n|^2}$$

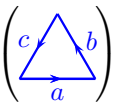
is the total quantum dimension (See Eq. 17.6). In Eq. 21.1, W is a weight assigned to each labeling of all the edges[8]. We consider the following definition of a weight assiggned to a given labeling of edges

$$W(\text{labeling}) = \frac{\prod_{\text{tetrahedra}} \tilde{G}(\text{tetrahedron}) \prod_{\text{edges}} d_{\text{edge}}}{\prod_{\text{triangles}} \tilde{\Theta}(\text{triangle})} \tag{21.2}$$

Thus each tetrahedron is given a weight $\tilde{G}$, depending on its labeling, each edge labeled a is given a weight d_a and each triangle is given a weight $\tilde{\Theta}^{-1}$ depending on its labeling.

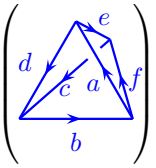
The weights $\tilde{G}$ and $\tilde{\Theta}$ are very closely related to quantities G and Θ we have already studied[9] in chapter 16 for example[10]. The functions $\tilde{G}$ and $\tilde{\Theta}$ are given by[11]

$$\tilde{\Theta}\left(\quad\right) = \Theta(a,b,c) = \sqrt{d_a d_b d_c} \tag{21.3}$$



and

$$\tilde{G}\left(\quad\right) = G^{bad}_{ecf} = F^{bad}_{ecf}\, d_f \sqrt{\frac{d_b d_c}{d_f}} \sqrt{\frac{d_a d_e}{d_f}} \tag{21.4}$$



Note that the tetrahedron shown here is different from the one shown in Fig. 16.14 that defines G from a planar diagram (or perhaps more properly a diagram drawn on the surface of a sphere). In fact the two tetrahedra are *dual* to each other. For example, in Fig. 16.14 the lines $f, e, \bar{c}$ form a loop whereas $f, \bar{e}, \bar{a}$ meet at a point. In the diagram in Eq. 21.4 on the other hand $e, f, \bar{c}$ meet at a point where $f, \bar{e}, \bar{a}$ form a loop. In Eq. 21.4 the three edges around any face must fuse together to the vacuum. I.e., we have the four conditions

$$N_{bad} > 0 \qquad N_{c\bar{d}e} > 0 \qquad N_{f\bar{e}\bar{a}} > 0 \qquad N_{\bar{c}\bar{f}\bar{b}} > 0$$

or else $\tilde{G}$ will vanish. Note that, like G, the value of $\tilde{G}$ is unchanged under any rotation of the tetrahedron.

21.2.1 Proof Turaev-Viro is a Manifold Invariant

The proof that $\mathcal{Z}_{TV}(\mathcal{M})$ is a manifold invariant is not difficult — one only needs to show that it is unchanged under the 1-4 and 2-3 Pachner moves. This is basically an exercise in careful bookkeeping (see exercise 21.2). Roughly, however, it is easy to see how it is going to work.

[8] In the language of statistical physics we can think of W as a Boltzmann weight for each edge label configuration, although it need not be positive, or even real.

[9] Many works, including the original works by Turaev and Viro [1992], use the diagrammatic algebra based on Temperly-Lieb which we discussed in chapter 19. However, in those works, they have used the nonunitary version of the diagrammatic algebra without the vertex renormalization which we introduce in section 19.3. In such an approach $\Theta(a,b,c)$ is replaced by $\Delta(a,b,c)$, for example (See Eq. 19.11). It is easy to show that these vertex renormalization factors completely cancel and the end value of the Turaev-Viro invariant is independent of whether the renormalization factors are included or not. Indeed, it is not necessary to have a fully unitary algebra for the Turaev-Viro construction to give a well behaved manifold invariant. We only need a consistent planar diagrammatic algebra. See also next margin note!

[10] In chapter 16 we insist on a fully isotopy invariant algebra with tetrahedral symmetry, and we will continue to assume those simplifications here. However, for constructing a Turaev-Viro invariant it turns out to be sufficient to have a spherical (hence pivotal) tensor category as we discuss in chapter 12. Full isotopy invariance is not required. This is discussed in depth by Barrett and Westbury [1996].

[11] See the comments in chapter 16 about how to choose the signs of the square-roots in cases where some d's are chosen negative.

Let us first examine the 2-3 Pachner move shown in Fig. 21.5. On the left we have two tetrahedra (call them 1 and 2) which are joined along a triangle (call it α). On the right we have three tetrahedra (call them 3, 4 and 5 which are joined along three triangles (call them β, γ, and δ) with the three triangles intersecting along a new edge down the middle (shown vertical in the figure) which we label with the quantum number n. To show that the Z_{TV} remains invariant we need to show that

$$\tilde{G}(1)\tilde{G}(2)\tilde{\Theta}(\alpha) = \sum_n \tilde{G}(3)\tilde{G}(4)\tilde{G}(5)\tilde{\Theta}(\beta)\tilde{\Theta}(\gamma)\tilde{\Theta}(\delta)d_n$$

The factors of $\tilde{\Theta}$ are simply factors of $\sqrt{d_a}$ and these cancel some factors of $\sqrt{d_a}$ in the definition of $\tilde{G}$ in Eq. 21.4. After this cancellation what remains is a relationship between two F's on the left and a sum over three F's on the right. The relationship that remains is exactly the pentagon equation Eq. 16.3 (or Eq. 9.7)! Thus any diagrammatic algebra which satisfies the pentagon equation will result in a Turaev-Viro partition function (Eq. 21.1) that is invariant under the 2-3 Pachner move!

The case of the 1-4 Pachner move is only a bit harder and we will sketch the calculation here. The large tetrahedra on the left of Fig. 21.4 (lets call this large tetrahedron 1) needs to be equivalent to the four smaller tetrahedra on the right (lets call these small tetrahedra $3, 4, 5$ and 6) once we sum over the quantum numbers on the four internal edges on the right. The three tetrahedra $3, 4$ and 5 share a common edge, and this is entirely analogously to the three tetrahedra we considered in the case of the 2-3 Pachner move. Summing over the quantum number of this common edge, and using the same pentagon relation replaces the three tetrahedra $3, 4, 5$ with two tetrahedra 1 and 2, where 1 is the large tetrahedron and 2 includes exactly the same edges as the remaining small tetrahedron 6. The tetrahedra 2 and 6 have 3 edges which are not shared with tetrahedron 1 — these are the remaining internal edges that need to be summed over. Summing over one of these internal edges, one invokes the consistency condition Eq. 16.5 to create a delta function which then kills one of the two remaining sums. The last remaining sum just yields a factor of $\mathcal{D}^2 = \sum_n d_n^2$ which accounts for the prefactor in Eq. 21.1 being that we have removed one vertex from the lattice.

21.2.2 Some TQFT Properties

The Turaev-Viro state sum has all the properties we expect of a TQFT. Although we need to discretize our manifold, the resulting "partition function" $Z_{TV}(\mathcal{M})$ for a manifold $\mathcal{M}$ is a complex number which is indeed independent of the discretization and depends on the topology of the manifold only.

As we discuss at length in section 7.1 we would also like $Z_{TV}(\mathcal{M})$ to represent a *wavefunction* if $\mathcal{M}$ is a manifold with boundary. To remind the reader, the point of this construction is that when we glue together two manifolds with boundary to get a closed manifold, this corresponds to taking the inner product between the two corresponding

wavefunctions to get a complex number.

To see how this occurs let us consider discretizing a manifold with a boundary. Here the 3D bulk of the manifold $\mathcal{M}$ should be discretized into into tetrahedra, and the 2D boundary surface $\Sigma = \partial M$ should be discretized into triangles. We divide the edge degrees of freedom into bulk and boundary where a boundary edge is defined as an edge where both vertices are on the boundary and all other edges are defined to be bulk. We define $Z(\mathcal{M})$ of such a discretized manifold with boundary as a sum like Eq. 21.1 where the sum is only over the edges in the bulk, leaving fixed (un-summed) the quantum numbers for the edges that live entirely on the boundary (i.e., both vertices on the boundary). Thus for manifolds with boundary we more generally write

$$Z_{TV}(\mathcal{M}; a_1, \ldots, a_N) = \mathcal{D}^{-2N_v - n_v} W'(a, \ldots, a_N) \sum_{\text{bulk labelings}} W(\text{bulk labels})$$

where N_v is the number of vertices in the bulk and n_v the number of vertices on the boundary. The weight function W is exactly the same as the weight function in Eq. 21.2 but only including edges, triangles, and tetrahedra in the bulk (All tetrahedra are considered bulk, and a triangle is considered boundary only when all three vertices are on the boundary). Here $a_1, \ldots, a_N$ are the quantum numbers of the edges on the boundary, and these are not included in the sum over bulk lables. An additional weight is included which is a function of these boundary edge labels

$$W'(a_1, \ldots, a_N) = \sqrt{\frac{\prod_{\text{boundary edges}} d_{\text{edge}}}{\prod_{\text{boundary triangles}} \tilde{\Theta}(\text{triangle})}}$$

The partition function $Z_{TV}(\mathcal{M}; a_1, \ldots, a_N)$ is now a function of the edge variables and is interpreted as a wavefunction[12] $|Z(\mathcal{M})\rangle$ that lives on the boundary $\Sigma = \partial M$.

[12]The wavefunction here takes some complex scalar value as a function of the physical variables which are the quantum numbers on the edge.

It is then quite natural to see how two manifolds can be glued together along a common boundary as in Fig. 7.3. In that figure we have a closed manifold $\mathcal{M} \cup_\Sigma \mathcal{M}'$ where $\mathcal{M}$ and $\mathcal{M}'$ are manifolds with boundary joined along their common boundary $\Sigma = \partial\mathcal{M} = [\partial\mathcal{M}']^*$. When we glue together $\mathcal{M}$ and $\mathcal{M}'$ we obtain the partition function for the full manifold as in Eq. 7.1 where we obtain the inner product by summing over the degrees of freedom of the wavefunction — which in this case means summing over the quantum numbers $a_1, \ldots, a_N$ of the edges on the boundaries. In other words, we have

$$\begin{aligned} Z_{TV}(\mathcal{M} \cup_\Sigma \mathcal{M}') &= \langle Z_{TV}(\mathcal{M}') | Z_{TV}(\mathcal{M}) \rangle \\ &= \sum_{a_1, \ldots, a_N} [Z_{TV}(\mathcal{M}'; \bar{a}_1, \ldots, \bar{a}_N)]^* \; Z_{TV}(\mathcal{M}; a_1, \ldots, a_N) \quad (21.5) \\ &= \sum_{j_1, \ldots, j_N} Z_{TV}(\mathcal{M}'; a_1, \ldots, a_N) \; Z_{TV}(\mathcal{M}; a_1, \ldots, a_N) \quad (21.6) \end{aligned}$$

where in the second line the edge variables in the first term are inverted because the surface of $\mathcal{M}'$ has the opposite orientation from the surface of $\mathcal{M}$. Going from the second to third line is an easy exercise (See exercise 21.1). The final result is easily see to be the correct expression for the Turaev-Viro invariant for the full manifold $\mathcal{M} \cup \mathcal{M}'$. I.e., it now sums over all the quantum numbers in both bulks and on the common boundary.

As in section 7.2 on can generalize the idea of a TQFT to include particle world lines (labeled links) as a well as the space-time manifold $\mathcal{M}$. As mentioned there we can roughly think of these world lines as internal boundaries, and we just fix the quantum number of edges along these hollow tubes to describe different world-line types. (See references at the end of the chapter)

21.3 Connections to Quantum Gravity Revisited

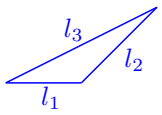
The Turaev-Viro invariant is a natural descendant of one of the very earliest approaches to quantum gravity pioneered by Penrose [1971] and Ponzano and Regge [1968]. Indeed much of the continued interest in Turaev-Viro and similar state-sum invariants is due to this relationship.

An interesting approach to macroscopic general relativity, used for example, in numerical simulation, is to discretize space-time into simplices — tetrahedra in three dimensions or four dimensional simplices (sometimes known as pentachora) in four dimensions[13]. The curvature of the space-time manifold (the metric) is then determined by the lengths assigned to the edges[14].

If one then turns to quantum gravity, one wants to follow the Feynman prescription and perform a sum over all possible metrics as we discussed previously in section 6.1. We can write a quantum partition function as

$$Z = \int \mathcal{D}g \; e^{iS_{Einstein}[g]/\hbar} \tag{21.7}$$

We can imaging performing such as sum for a discretized system by integrating over all possible lengths of all possible edges. However, not all triangle edge length should be allowed — in Euclidean space one must obey the crucial constraint of the triangle inequality[15]



$$\Rightarrow \qquad |l_1 - l_2| \leq l_3 \leq (l_1 + l_2) \tag{21.8}$$

The key observation is that the triangle inequality is precisely the same as the required inequality for regular angular momentum addition

$$j_1 \otimes j_2 = |j_1 - j_2| \oplus |j_1 - j_2| + 1 \oplus \ldots \oplus |j_1 + j_2|. \tag{21.9}$$

Thus it is natural to label each edge of with a quantum mechanical spin,

[13] It is also possible to discretize space and leave time continuous. This leaves some concerns with Lorentz invariance but may have other advantages. Other discretization approaches also exist, see Regge and Williams [2000].

[14] All of general relativity can be reformulated in this discrete language. This is known as *Regge calculus*. See Regge [1961].

[15] These inequalities must hold even with a curved spatial metric.

and sum over all possible spins. Such an approach is known as a *spin network*. We thus imagine building a Turaev-Viro model (Eq. 21.1) with a planar diagrammatic algebra built from angular momentum addition rules: quantum numbers are the angular momenta j, the fusion rules are as given in Eq. 21.9, and the F-matrices are given by the regular $6j$ symbols of angular momenta addition[16]. Such a model turns out to be very precisely[17] the quantum gravity partition function Eq. 21.7 (up to the fact that one still needs an additional sum over topologies of the space-time manifold if one wants a full sum over all possible histories)! As we expect from the discussion in chapter 6, the resulting description of quantum gravity in 2+1D is a TQFT.

There is, unfortunately, one clear problem with this approach. Because there are an infinite number of different representations of $SU(2)$—i.e., an infinite number of different values for the angular momentum quantum number j — the partition function sum formally diverges. This divergence becomes regularized if we find a way to consistently cut off the sum over angular momenta at some maximum value k. Using the diagrammatic rules of $SU(2)_k$ (the same diagrammatic rules we built up in chapter 19, see in particular margin note 6) implements this cutoff and yields a divergence-free result[18].

21.4 Dijkgraaf-Witten Model

Another state sum model of some interest is the so-called Dijkgraaf-Witten model[19] (Dijkgraaf and Witten [1990]). As with Turaev-Viro this model discretizes space into simplices and sums over possible labels of all the edges.

In the Dijkgraaf-Witten model we choose a group G and we label the edges of the simplices with elements from that group. The general idea is very similar to that of Turaev-Viro just using the multiplication properties of the group to give us a set of fusion rules as in section 20.1 and we use a 3-cocycle in place of the F-matrix[20]. These fusion rules require that multiplication of the group elements around every triangle must result in the identity as shown in Fig. 21.6. This is the analog of Eq. 21.3 where three quantum numbers around a triangle must fuse to the identity. This condition is known as a "flatness" condition, with the name coming from lattice gauge theory, which we will see in more detail in chapter ***.

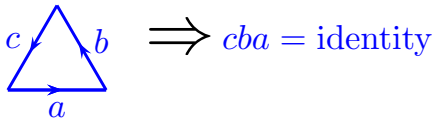


Fig. 21.6 Multiplying group elements around a triangle in Dijkgraaf-Witten theory results in the identity. This is known as the "flatness" condition

[16] Building a diagrammatic algebra based on a Lie group ($SU(2)$ in this case) is mentioned in section 20.2.3 above.

[17] Although the idea of spin networks as a toy model for quantum gravity goes back to Penrose [1971], and was pursued further by Ponzano and Regge [1968], it was only much later that Hasslacher and Perry [1981] showed a more precise equivalence of the model to gravity.

[18] As we will see in section *** below, the Turaev-Viro model built from the $SU(2)_k$ diagrammatic rules is equivalent to the so-called *quantum-double* Chern-Simons theory $SU(2)_k \otimes SU(2)_{-k}$. As we mentioned in section 6.3 above, such a Chern-Simons theory is equivalent to 2+1D gravity with a cosmological constant $\lambda = (4\pi/k)^2$. Taking the limit of large k then gives the classical limit of simple $SU(2)$ angular momentum addition corresponding to a universe with no cosmological constant.

[19] Robbert Dijkgraaf is a very prominent theoretical physicist and string theorist. His surname is likely to be difficult to properly pronounce for those who are not from the Netherlands because the "g" is a gutteral sound that only exists in Dutch. However, those from the south of the Netherlands don't use the gutteral "g" and instead pronunce it as Dike-Hraff, which is probably about the closest most English speakers will get to the right result. The word "Dijkgraaf" refers to an occupation: A Dijkgraaf is the person in charge of making sure that water stays in the ocean and does not flood the cities and the rest of the Netherlands.

[20] For the case of an abelian group Dijkgraaf-Witten is a special case of Turaev-Viro. However Turaev-Viro does not consider fusion rules where $g \times h = h \times g$ so for nonabelian groups Dijkgraaf-Witten is not just a special case of Turaev-Viro. The group need not be abelian since we only need to have an algebra that is consistent on a plane (or sphere) in order to defne its value on a tetrahedron (see the comments in section 20.1.3).

As mentioned in section 20.1 when we use group multiplication for fusion rules, the quantum dimensions[21] of all the particles are all $d_a = 1$. This means that in Eq. 21.2 both the d_a factor and the $\tilde{\Theta}$ factor are trivial. We are thus left with only the tetrahedron factor and the Dijkgraaf-Witten partition function looks like a simplified version of the Turaev-Viro case in Eqs. 21.1 and 21.2 given by[22]

$$Z_{DW}(\mathcal{M}) = |G|^{-N_v} \sum_{\text{labelings}} \prod_{\text{tetrahedra}} \tilde{G}(\text{tetrahedron}) \tag{21.10}$$

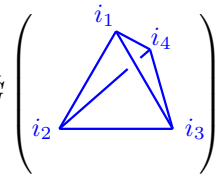
where N_v is the number of vertices, $|G|$ is the number of elements in the group G, and the sum is only over labellings that satisfy the flatness condition (Fig. 21.6).

The tetrahedral symbol $\tilde{G}$ is a bit more complicated than in the case of the Turaev-Viro invariant. We do not generally have full tetrahedral symmetry so it could matter which way we orient the tetrahedron when we evaluate $\tilde{G}$. In order to define the tetrahedral symbol $\tilde{G}$ properly we do the following: First we label each vertex in the system with a unique integer (it will not matter which vertex gets which label!). Given a tetrahedron with vertices i_1, i_2, i_3, i_4 we sort these vertices in ascending order so that

$$[j_1, j_2, j_3, j_4] = \text{sort}[i_1, i_2, i_3, i_4] \quad \text{such that} \quad j_1 < j_2 < j_3 < j_4$$

we then define

$$\tilde{G}\left(\quad \right) = \omega(g_{j_2,j_1}\ ,\ g_{j_3,j_2}\ ,\ g_{g_4,g_3})^{s(j_1,j_2,j_3,j_4)} \tag{21.11}$$



Here $g_{k,l}$ is the group element on the edge directed from vertex k to vertex l, and ω is the chosen 3-cocycle. The exponent $s(j_1, j_2, j_3, j_4)$ is either $+1$ or -1 depending on whether the orientation of the tetrahedron defined by the ordered set of vertices $[j_1, j_2, j_3, j_4]$ has the same or opposite orientation as the manifold we are decomposing[23]. This prescription gives a manifold invariant (The Dijkgraaf-Witten invariant) for any choice of 3-cocycle even if the corresponding diagrammatic algebra does not have isotopy invariance. (See exercise ***).

21.4.1 Other Dimensions

An interesting feature of Dijkgraaf-Witten theory is that essentially the same recipe builds a Dijkgraaf-Witten TQFT in any number of dimensions. One discretizes the D-dimensional manifold into D-dimensional simplices (segments in 1D, triangles in 2D, tetrahedra in 3D, pentachora in 4D) and labels each edge with a group element $g \in G$ and each vertex is assigned an integer label. The flatness condition is always the same as that shown in Fig. 21.1 — multiplying the group ele-

[21] In chapter 20 we considered also the possibily of $d_a = -1$ but this is a gauge choice. We are always entitled to chose $+1$ instead at the cost of possibly losing isotopy invariance.

[22] With apologies for using G and $\tilde{G}$ in the same equation to mean completely different things!

[23] To find the orientation of a tetrahedron, place j_1 closest to you and see if the triangle $[j_2, j_3, j_4]$ is oriented clockwise or counterclockwise.

ments around a closed loop must give the identity. In D-dimensions we build the partition function by multiplying a weight for each D-simplex, where the weight is given now by a so-called D-cocycle[24] which we call $\omega_D(g_1, g_2, \ldots, g_D)$ which is now a function of D arguments. Finally, one builds a partition function by summing over all possible labelings

$$Z_{DW}(\mathcal{M}_D) = |G|^{-N_v} \sum_{\text{labelings}} \prod_{\text{D-simplices}} \omega_D(g_{j_2,j_1}, \ldots, g_{j_{D+1},j_D})^{s(j_1,\ldots,j_{D+1})} \tag{21.12}$$

As with the 3D case, the arguments of the cocycle $g_{k,l}$ are the group elements along the edges of the simplex from vertex k to vertex l and we always write them ordered such that $j_1 < j_2 < \ldots < j_D$. Finally the exponent s is always ± 1 depending on whether the orientation of simplex described by the ordered set $[j_1, \ldots, j_{D+1}]$ matches that of the underlying manifold or not.

As a quick example, let us consider the 2D case. The definition of a 2-cocycle ω_2 is any function that satisfies the condition[25]

$$\omega_2(g,h)\omega_2(gh,k) = \omega_2(h,k)\omega_2(g,hk) \tag{21.13}$$

In the partition function, Eq. 21.12, each triangle gets a weight given by the cocycle. It is then easy to see that the cocycle condition is precisely the condition necessary to make the partition function invariant under the 2-2 Pachner move, as shown in Fig. 21.7. It is not a hard exercise to demonstrate invariance under the 3-1 Pachner move as well (See exercise 21.3).

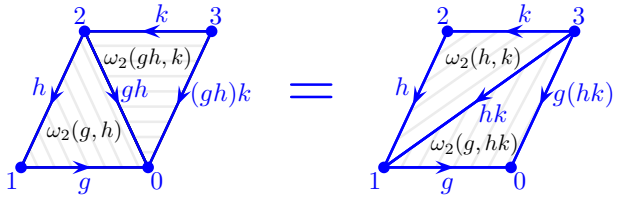


Fig. 21.7 Each triangle satisfy the flatness condition Eq. 21.6 meaning multiplying all three edges in order gives the identity. In the partition function each triangle gets a weight given by the corresponding cocycle ω_2 as written in black text. All of the triangles in the figure are oriented positively $s = +1$. The cocycle condition Eq. 21.13 guarantees that the product of the cocycles on the left equals the product of the cocycle on the right.

21.4.2 Further Comments

One particularly interesting special case of Dijkgraaf-Witten theory is the case of the trivial 3-cocycle where ω is always unity. In this case, the argument of the sum in Eq. 21.10 (or more generally Eq. 21.12) is just unity so the partition function just counts the number of flat field configurations (See Fig. 21.6) and then divides by $|G|^{N_v}$. This partition function is exactly that of lattice gauge theory, as we will see in chapter *** below, and the resulting topological quantum field theory is known

[24] I won't give the most general definition of cocycle as this takes us too far afield into group cohomology. However, as with the 3-cocycle it is simply a function satisfying a particular cocycle condition. See Eq. 20.1 for the 3D case and Eq. 21.13 for the 2D case.

[25] The 2-cocycle condition is equivalent to the consistency condition for a so-called "projective representation" of the group. For projective representations we have the multiplication rule $\rho(g)\rho(h) = \omega_2(g,h)\rho(gh)$ whereas for regular group representations we have $\omega_2 = 1$. See section 33.2.4.

as the quantum double of the group G. The more general case, with a nontrivial cocycle is correspondingly sometimes known as "twisted" gauge theory, where the cocycle is thought of as some sort of twist to the otherwise simple theory.

A further interesting relatioship is that Dijgraaf-Witten theory can be thought of as result of symmetry breaking an appropriately chosen Chern-Simons theory (See for example Dijkgraaf and Witten [1990], de Wild Propitius [1995]). One might imagine, for example, breaking a compact $U(1)$ Chern-Simons gauge theory into a discrete $\mathbb{Z}_n$ group — like breaking the symmetry of a circle into an n-sided regular polygon. The particular cocycle one gets in the resulting Dijkgraaf-Witten theory depends on the choice of the coefficient (the "level") of the Chern-Simons term.

Dijkgraaf-Witten theory has had extensive recent applications within quantum condensed matter physics where it turns out that a classification of so-called symmetry protected topological (SPT) phases is given in terms of Dijkgraaf-Witten theories. We will briefly discuss SPT phases in section *** below.

Further Reading

- The Turaev-Viro invariant was introduced in Turaev and Viro [1992]. Rather interestingly Turaev and Viro were apparently unaware of the earlier work by Penrose, Ponzano, Regge and others when they first discussed these state sums! The work was extended to include all spherical fusion categories by Barrett and Westbury [1996]. A recent rather complete discussion, including looking at the possible world-lines and boundaries is given by in the book Turaev and Virelizier [2017]. Unfortunately, these references and many other works in the field are written in rather mathematical language that is not particularly transparent for most physicists!
- It is worth commenting that the state-sum approach to quantum gravity has been extended in a multitude of ways, and continues to be an active area of research. Among the key directions are extension to 3+1 dimensions (Ooguri [1992] and Crane and Yetter [1993] for example), and extensions to Lorentzian signature (Barrett and Crane [2000]). A nice general dicussion of discrete approaches gravity is given by Regge and Williams [2000].
- One very popular extension of the spin-network modes, known as a *spin-foam*, is to discretize space but allow the discretization to change as a function of time. A nice review of this direction is given by Lorente [2006].

Exercises

Exercise 21.1 Some More Facts about Turaev-Viro

Consider a manifold $\mathcal{M}$ with boundary Σ which has been discretized into tetrahedra on in the bulk and triangles on the surface. Let the edges on the surface be labeled by $j_1, \ldots, j_N$. Assume tha the theory has relflection symmetry as in Eq. 16.14, show that

$$[Z_{TV}(\mathcal{M}; \bar{a}_1, \ldots, \bar{a}_N)]^* = Z_{TV}(\mathcal{M}; a_1, \ldots, a_N)$$

And as a result show that for a closed manifold $Z(\mathcal{M})$ is real.

Exercise 21.2 Details of Turaev-Viro

Work carefully through the details of the proof that the Turaev-Viro partition function is invariant under Pachner moves.

Exercise 21.3 2D Dijkgraaf-Witten

The invariance of the two dimensional Dijkgraaf-Witten partition function under the 2-2 Pachner move is established in section 21.4.1. Show that the partition function is also invariant under the 3-1 Pachner move.

Formal Construction of TQFTs from Diagrams: Surgery and More Complicated 3-Manifolds[1]

22

Having constructed diagrammatic algebras in 2+1 dimension[2], we have almost all we need to define a TQFT based on these diagrams. As discussed in section 14.3 our diagrammatic algebra which gave us a way to evaluate a partition function Z(labeled link in S^3)/$Z(S^3)$, or equivalently Z(labeled link in $S^2 \times S^1$) with the caveat that no link goes around the handle of S^1. However, a TQFT should be able to evaluate a partition function in *any* arbitrary manifold $\mathcal{M}$. Indeed, in the simplest case we might dispense with the labeled link and want to find a partition function of the manifold $\mathcal{M}$ alone.

In this chapter we develop a prescription for handling more complicated manifolds. One important thing this will achieve will be to give a formal definition to Chern-Simons theory, which we like to think of as being defined as some sort of functional integral, but as pointed out in section 5.3.4 is not really well defined in that language as such integrals do not actually converge.

The way we will handle more complicated manifolds is by sewing pieces of manifolds together with a procedure known as surgery.

[1] Although this chapter is super interesting and fun, physicists can probably skip it on a first reading.

[2] I.e., including braiding again, which we did not need to define the TQFTs in chapter 21.

22.1 Surgery

In chapter 7 we saw two examples of assembling manifolds by gluing together pieces. We found that we could assemble together two solid tori ($D^2 \times S^1$) into either S^3 or $S^2 \times S^1$ depending on how we glue together the $S^1 \times S^1$ surfaces. (In fact, one can consider gluing together the surfaces in yet other ways to get even more interesting results[3], but we will not need that for the moment). We would like to use this sort of trick to study much more complicated three dimensional manifolds.

The understanding of three dimensional manifolds is a very rich and beautiful problem[4]. In order to describe complicated manifolds it is useful to think in terms of so-called surgery. Similar to what we were

[3] See for example the discussion in section 7.4 as well as Rolfson [1976] for example.

[4] Many important results on three dimensional manifolds have been discovered recently. Perelman's[5] proof of the Poincaré Conjecture, along with the methods he used are apparently extremely revolutionary and powerful. But this is *way* outside the scope of our book!

[5] Grigori Perelman is a brilliant, but startlingly puzzling character. He famously declined the million dollar Millenium Prize offered to him for proving the Poincaré conjecture in three dimensions. He turned down the Fields Medal as well.

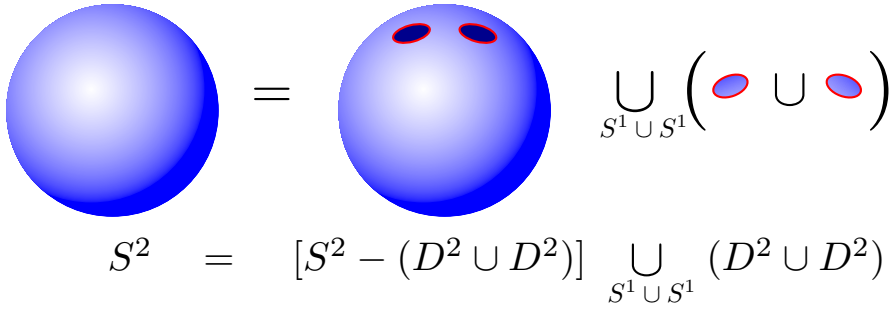


Fig. 22.1 Writing a sphere $\mathcal{M} = S^2$ as the union of two manifolds glued along their boundaries. $\mathcal{M}_2$ is the union of two disks $D^2 \cup D^2$. $\mathcal{M}_1 = S^2 - (D^2 \cup D^2)$ is the remainder. The two manifolds are glued along their common boundary $S^1 \cup S^1$.

discussing in section 7.3 — assembling a manifold by gluing pieces together — the idea of surgery is that we remove a part of a manifold and we glue back in something different. Imagine replacing someone's foot with a hand![6] By using successive surgeries we will be able to construct any three-dimensional manifold[7].

The general scheme of surgery is to first write a manifold as the union of two manifolds-with-boundary sewed along their common boundaries. If we have a closed manifold $\mathcal{M}$ that we would like to alter, we first split it 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 $\mathcal{M}'$. In other words, we have simply thrown out the $\mathcal{M}_2$ part of the manifold and replaced it with $\mathcal{M}_2'$.

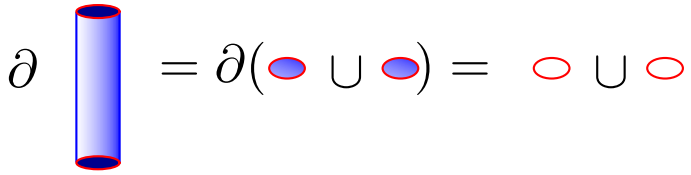
[6] Prehensile toes could be useful I suppose!

[7] We will only be concerned with orientable manifolds

22.1.1 Simple Example of Surgery on a 2-manifold

To give an example of surgery consider the sphere $\mathcal{M} = S^2$ as shown in Fig. 22.1. Here we write the sphere as the union of two disks $\mathcal{M}_2 = D^2 \cup D^2$ and the remainder of the sphere $\mathcal{M}_1 = S^2 - (D^2 \cup D^2)$. These are glued along their common boundary $S^1 \cup S^1$.

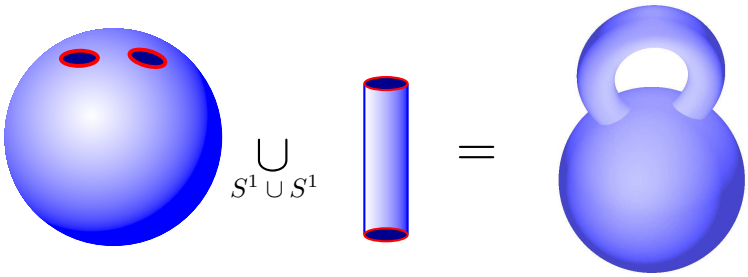
Now we ask the question of what other 2-manifolds have the same boundary $S^1 \cup S^1$. There is a very obvious one, the cylinder surface! Let us choose the cylinder surface $\mathcal{M}_2' = S^1 \times I$ where I is the interval (or D^1). It also has boundary $\partial\mathcal{M}_2' = S^1 \cup S^1$ as shown in Fig. 22.2.



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

Fig. 22.2 The boundaries of the cylinder surface is the same as the boundary of the two disks. Both boundaries are two circles. This means that we can remove two disks from a manifold and sew in the cylinder.

Thus we can sew the cylinder surface in place where we removed $\mathcal{M}_2 = D^2 \cup D^2$, as shown in Fig. 22.3. The resulting manifold $\mathcal{M}'$ is the torus T^2



$$[S^2 - (D^2 \cup D^2)] \underset{S^1 \cup S^1}{\cup} [S^1 \times I] = T^2$$

Fig. 22.3 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. Note that the object on the right is topologically a torus.

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. 22.4

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

Fig. 22.4 Handle attaching on the manifold $\mathcal{N}$ is the same as surgery on a manifold $\mathcal{M} = \partial\mathcal{N}$.

22.1.2 Surgery on 3-manifolds

We can also perform surgery on three-dimensional manifolds[8]. Start with a simple closed 3-manifold $\mathcal{M}$, such as S^3 (or, even simpler to think about, consider $\mathcal{M} = \mathbb{R}^3$ and let us not worry about the point at infinity). 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 solid torus with exactly the same surface:

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

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[9]. The reason this is difficult to visualize is because if we start with a very simple space like $\mathcal{M} = \mathbb{R}^3$ the new structure $\mathcal{M}'$ is not embeddable within the original manifold $\mathcal{M}$.

This procedure, torus surgery on a 3-manifold,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, see section 7.4) 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. As long as we have a closed loop S^1 (possibly following a complicated path) and it is thickened to D^2 in the direction transverse to the S^1 path, it is still a solid-torus topologically.

[8] This is the part that is guaranteed to make your head explode.

[9] Stop here, think about what we have done. Collect the pieces of your exploded head.

22.2 Representing Manifolds with Knots

22.2.1 Lickorish-Wallace Theorem

An important theorem[10] of topology is due to Lickorish [1962] and Wallace [1960].

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

One has the following procedure. We start with a link (some knot

[10] In Witten's groundbreaking paper on the Jones polynomial (Witten [1989]), he states the theorem without citation and just says "It is a not too deep result...". Ha!

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[11]. Any possible 3-manifold can be obtained in this way by surgering an appropriately chosen link. We summarize with the mapping

$$\text{Link in } S^3 \overset{\text{surger}}{\longrightarrow} \text{Some } \mathcal{M}^3 \tag{22.1}$$

We can thus represent any three dimensional manifold as a link in S^3. 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 exploit this connection further when we discuss the Witten-Reshitikhin-Turaev invariant below in section 22.3.

[11] See also section 7.4.

[12] If one does not start with the knot embedded in S^3, one may need a third move known as "circumcision". This says that if any string loops only once around another string (without twisting around itself and without looping around anything else), both strings may be removed. I.e., in Fig. 22.5, both strings may be removed (independent of how the string going off to the left forms any knot). See exercise **??**.

22.2.2 Kirby Calculus

It turns out that not all topologically different links, when surgered, give topologically different manifolds. Fortunately, the rules for which knots give the same manifolds have been worked out by Kirby [1978]. These rules, known as Kirby calculus, are stated as a set of transformation moves 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 which can be combined together to make more complicated transformations. Perhaps the simplest set of two elementary basic moves are known as Kirby moves which we will present here[12]. We will not rigorously prove that these moves leave the manifold unchanged, but we will give rough arguments instead.

Fig. 22.5 A circumcision. Both strings can be removed. This is a third Kirby move which is implied by the first two if you start with a link embedded in S^3 but is more generally an independent move that is required. See for example Roberts [1997].

Kirby Move 1: Blow Up/ Blow Down:[13]

One can add or remove a loop with a single twist, as shown in Fig. 22.6, to a link and the manifold resulting after surgery remains unchanged.

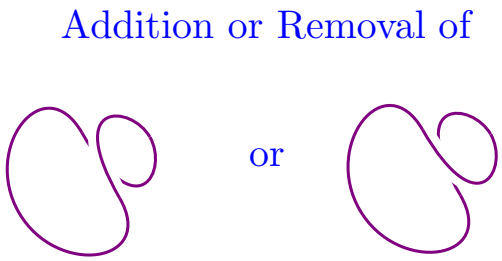


Fig. 22.6 Blow up/ Blow down. Addition or removal of an unlinked loop with a single twist leaves the 3-manifold represented by surgery on the knot unchanged.

[13] The nomenclature is obscure when discussing 3-manifolds, but makes sense when one discusses 4-manifolds. See any of the books on 4-manifold topology listed at the end of the chapter.

Argument: First let us be a bit more precise about the surgery prescription. Given a link, we think of this link as being a ribbon (usually

we draw it with blackboard framing, see section 15). Thicken each strand into a solid torus, and draw a line around the surface of this torus that follows one of the edges of the ribbon. Remove this solid torus, but the torus surface that remains still has the line drawn around it. Reattach a new solid torus where the new meridian (the circle surrounding the contractable direction) follows precisely this line.

Now consider a twisted loop as in Fig. 22.6 embedded in S^3. As shown in Fig. 2.7 a string with a small twist loop as in Fig. 22.6 can be thought of as a ribbon with a twist (but no loop) in it. Let us use this description instead. Thicken the loop to a torus, and then the ribbon traces out a line as shown in Fig. 22.7 on the torus surface. We remove the solid torus and insert a new torus where the meridian follows the twisted line on the surface of the hole that is left behind. This is exactly the construction of $L(1,1) = S^3$ described in 7.4 above (it is $(1,1)$ since the blue line goes around each handle once), thus showing an example of how surgery on the twisted loops in Fig. 22.6 does nothing to the manifold.

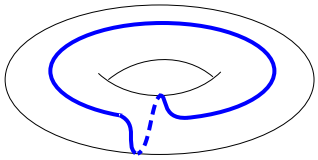


Fig. 22.7 A line that wraps both the longitude and meridian of the torus. If we thicken the knot shown in Fig. 22.6 to a torus and draw a line around the longitude of the torus, then try to straighten the torus out to remove the twist, the straight line ends up looking like this.

Kirby Move 2: Handle-Slide:[14]

A string can be broken open and pulled along the full path of another string, and then reconnected, and the resulting manifold remains unchanged. See Fig. 22.8 or 22.9.

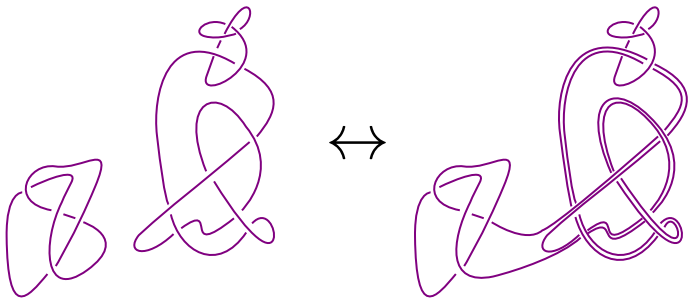


Fig. 22.8 A handle-slide move. (See Fig. 22.9 for another example.) Both left and right sides of this picture represent the same 3-manifold after surgery. Note that we should always view both strings as ribbons, and we need to keep track of how many self-twists the ribbon accumulates when it is slid over another string.

Argument: Consider the simple handle-slide shown in Fig. 22.9. Let us think about what happens when we surger the horizontal loop. First we thicken the horizontal loop into a torus (as shown), then we exchange the contractable and non-contractable directions. In this procedure, the longitudinal direction (The long direction) of the torus is made into something contractable. This means (after surgery) we can pull the far left vertical line through this torus without touching the three vertical blue lines. Thus the right and left pictures must describe the same manifold. While it is a bit harder to argue generally, this principle remains true even if the torus is embedded in the manifold in a complicated way,

[14] The nomenclature "handle slide" comes from an interpretation of this move as sliding handles around on a manifold. Consider the example used in section 22.1.1 where we attached a handle to a ball and obtained a solid torus. We could also attach two handles and get a two-handled solid torus. Here it doesn't matter where the handles are attached to the sphere – they can be slid around. Indeed, they can even be slid over each other (where one handle attaches to some point on the other handle). It is the sliding of a handle over another handle which gives this move its name.

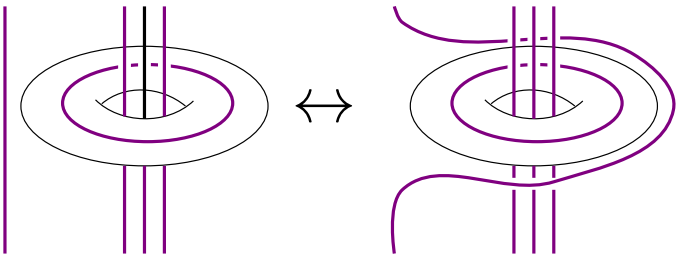


Fig. 22.9 An example of a simple handle-slide move.

as in Fig. 22.8.

Two links in S^3 describe the same 3-manifold if and only if one link can be turned into the other by a sequence of these Kirby moves as well as any smooth deformation of links (i.e., regular isotopy). Note that if we have two disconnected links L_1 and L_2 which surgered give two manifolds $\mathcal{M}_1$ and $\mathcal{M}_2$ respectively,

$$\begin{array}{ccc} L_1 & \overset{\text{surger}}{\longrightarrow} & \mathcal{M}_1 \\ L_2 & \overset{\text{surger}}{\longrightarrow} & \mathcal{M}_1 \end{array}$$

then if we consider a link $L_1 \cup L_2$ which is the disconnected union of the two links (i..e, the two links totally seperated from each other) it is fairly easy to see that we obtain the so-called connected sum[15] of the two manifolds which we write as follows:

$$L_1 \cup L_2 \quad \overset{\text{surger}}{\longrightarrow} \quad \mathcal{M}_1 \# \mathcal{M}_2$$

22.3 Witten-Reshitikhin-Turaev Invariant

By using the ideas of surgery, we are now in a position to use our diagrammatic algebra to handle complicated manifolds. Recall that one of the definitions of a TQFT is a mapping from a manifold $\mathcal{M}$ to a complex number $Z(\mathcal{M})$ in a way that depends only on the topology of the manifold (for example, Eq. 5.16 or Fig. 7.1 but without the embedded link). By using surgery (Eq. 22.1) we can describe our manifolds as links in S^3. If we can then find a link invariant that is unchanged under Kirby moves, we will effectively have something we can use as a manifold invariant. Thus we are now looking to construct a link invariant, and our diagrammatic algebra will provide exactly what we need!

We want to have a link invariant which is fully isotopy invariant (since Kirby calculus is isotopy invariant). In the simplest case let us assume we have no negative Frobenius-Schur indicators[16]so we can take all $d_a = \mathsf{d}_a$ positive and we have a fully isotopy invariant diagram algebra for links (as discussed in section 14.3).

The key to this construction is to consider a link of the Ω (Kirby)

[15]To form a connected sum of two d-dimensional manifolds, one deletes a small d-dimensional ball in each of the manifolds and sews together the two boundry spheres. For example in two dimensions the connected sum of two two-dimensional torus surfaces is a two-handled torus surface:

$$T^2 \# T^2 = \text{two handle torus}$$

.

strands discussed above in section 17.5.

This link made of Ω is represents the link to be surgered, and thus represents our manifold. Let us now consider a manifold invariant defined as

$$Z_{WRT}(\mathcal{M}) = \frac{1}{\mathcal{D}}\left[e^{2\pi ic/8}\right]^{\sigma}\left(\begin{array}{c}\text{Evaluate link made of } \Omega\text{-strands where}\\ \text{surgery on link in } S^3 \text{ gives } \mathcal{M}\end{array}\right) \tag{22.2}$$

where σ is the so-called signature of the link, defined to be the number of positive eigenvalues minus the number of negative eigenvalues of the matrix of linking numbers $link_{ij}$ between the (possibly multiple) strands of the link (The diagonal element $link_{ii}$ is just the self-linking or writhe of strand i. See section 2.6.2 for definition of linking number[17]).

[17]Note that to calculate a linking matrix, we must orient all of the strands (i.e., put arrows on them). It does not matter which way these arrows point.

It is not so obvious that the definition in Eq. 22.2 should provide a manifold invariant. What we would need to show is that $Z_{WRT}(\mathcal{M})$ gives the same output for *any* link that describes the same $\mathcal{M}$. In ther words we have to show that the expression on the right hand side of Eq. 22.2 is unchanged when we make Kirby moves on the link.

Let us consider the first Kirby move, the addition of a twisted loop as in Fig. 22.6. Using Fig. 17.9, the adding such a twisted loop multiplies the value of the link (the final term in Eq. 22.2) by $e^{\pm 2\pi ic/8}$ ($\pm$ depending on which way the loop is twisted). However, the addition of the twisted loop also changes the signature of the link σ by ∓ 1 thus precisely canceling this factor. Thus the expression in Eq. 22.2 is certainly unchanged under the first Kirby move[18], the Blow-Up/Blow-Down.

[18]The killing property of Fig. 17.7 also makes Eq. 22.2 invariant under the third Kirby move, Fig. 22.5.

We now turn to the second Kirby move. Here we show a rather remarkable property of the Ω-strand — it is invariant under handle slides! (up to phases which are properly corrected by the prefactor of Eq. 22.2). The derivation of this result is given in Fig. 22.10. One must be a bit cautious in applying this handleslide law, as the strand being slid (say the left strand in Fig. 22.10) can develop self-twists if it slides over a strand (say the right, Ω-strand in Fig. 22.10) which itself has twists. However, the phase prefactor of Eq. 22.2 is designed to precisely account for this. Thus Eq. 22.2 is unchanged under Kirby moves and therefore gives an invariant of the manifold.

The manifold invariant Eq. 22.2 is known as the Witten-Reshetikhin-Turaev invariant, and was invented by Reshetikhin and Turaev [1991]. The reason it also gets named after Witten is that it gives a rigorous redefinition of the Chern-Simons manifold invariants (Eq. 5.16) discussed by Witten [1989]. This is a rather important result being that the Chern-Simons functional integral is not well defined as an integral! (See the comments in section 5.3.4)[19].

[19]We mention in section 5.3.4 that the Chern-Simons partition function, among other reasons for being ill-defined, actually depends on a so-called 2-framing of the manifold. The Reshitikhin-Turaev invariant corresponds to choosing so-called *canonical* framing. This is discussed in depth by Atiyah [1990b]; Kirby and Melvin [1999].

[16]In the more general case where we have negative Frobenius-Schur indicators, we can use the scheme of section 14.5 to insure isotopy invarience. It is also possible to work work with negative d_a's as discussed in section 14.2.1 and *not* apply rule 0. In many discussions of Reshitikhin-Turaev invariant, this is essentially what is done. Note that in this case the Kirby color Fig. 17.6 is then defined with $d_a/\mathcal{D}$ rather than $\mathtt{d}_a/\mathcal{D}$. The identities of Fig. 17.7 and 17.9 still hold.

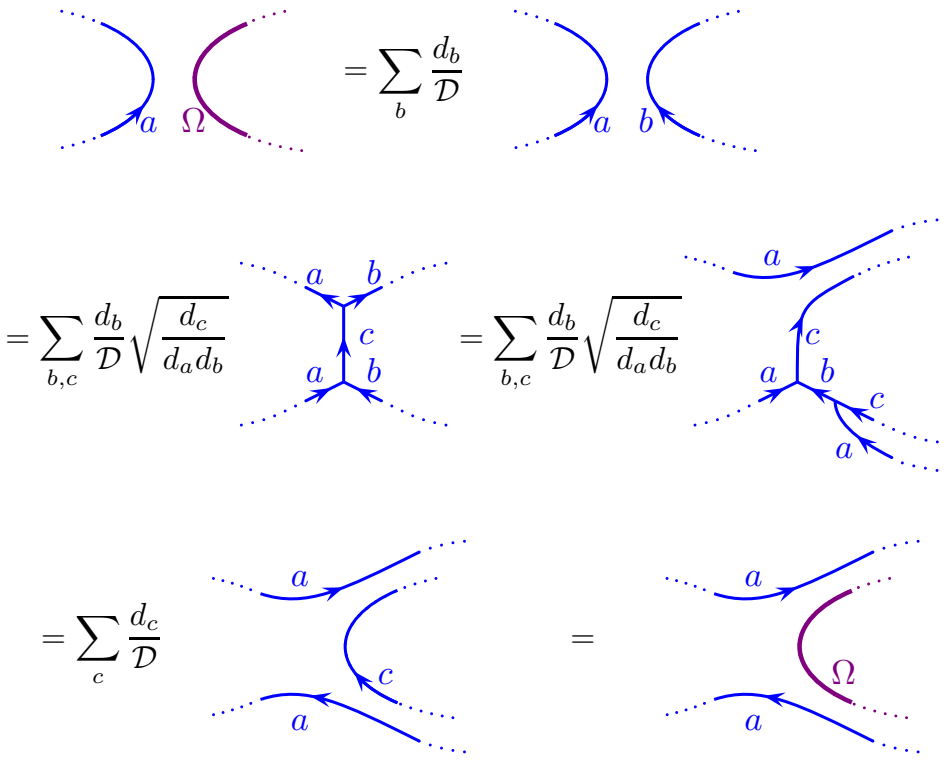


Fig. 22.10 Proof that the Ω strand satisfies the handle-slide. Here we show that any strand a can freely slide over the Ω strand in the sense of Fig. 22.8. The Ω strand on the right is meant to be connected up to itself in some way in a big (potentially knotted or linked) loop which we don't draw. In going from the first to the second line, and also in going from the second to the third line we have used the completness relation Eq. 16.8. The equality in the second line is just sliding the vertex from the top (where a and b split from c) all the way around the b strand on the right until it almost reaches the bottom a, b, c vertex. Note that if there are fusion multiplicites $N_{ab}^c > 1$ then there are additional sums over the vertex variables μ as in Eq. 16.24. However, this does not alter the result.

Note the multipication law for connected sums of manifolds

$$Z_{WRT}(\mathcal{M}_1 \# \mathcal{M}_2) = \frac{Z_{WRT}(\mathcal{M}_1) Z_{WRT}(\mathcal{M}_2)}{\mathcal{D}} \tag{22.3}$$

This multiplication law is from the fact that surgery on disjoint union of links gives a connect sum of manifolds (Eq. 22.2) and the evaluation of the disjoint union of links gives the product of the individual evaluation of the two links[20].

Further one can extend these manifold invariants to give a topological invariant partition function of a labeled link within a manifold as in Fig. 7.1 (this was one of our general definitions of what we expected

[20] Note that some references redefine Z_{WRT} without the factor of $1/\mathcal{D}$ out front such that $Z(\mathcal{M}_1 \# \mathcal{M}_2) = Z(\mathcal{M}_1) Z(\mathcal{M}_2)$ instead.

from a TQFT). To make this extension we simply define

$$Z_{WRT}(\mathcal{M};\text{labeled link}) = \frac{1}{\mathcal{D}}\left[e^{2\pi ic/8}\right]^{\sigma}\left(\begin{array}{c}\text{Evaluate link made of}\\ \Big(\Omega\text{-strands where}\\ \text{surgery on link in } S^3 \text{ gives } \mathcal{M}\\ \cup \text{ labeled link}\Big)\end{array}\right)$$

In other words, we simply include the labeled link into the diagram to be evaluated.

Without ever saying the words "path integral" or "Chern-Simons action" we 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), and surgery on Ω strands allows us to represent complicated manifolds.

22.3.1 Some examples

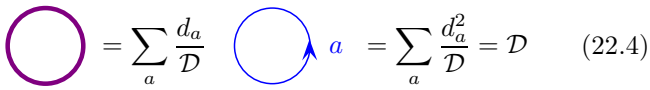
It is worth working through a few examples of calculating the Witten-Reshitikhin-Turaev Invariant for a few simple manifolds.

For $\mathcal{M} = S^3$, we don't need to surger the manifold at all, so don't need any Ω link at all. The value of the (empty) link is normalized to unity and including the prefactor in Eq. 22.2 (with signature zero) we obtain

$$Z_{WRT}(S^3) = 1/\mathcal{D}$$

which matches our expectations given Eqs. 17.9 and 7.7.

For $\mathcal{M} = S^2 \times S^1$ we need to surger a single loop in S^3 to obtain $S^2 \times S^1$ (See exercise 22.1). Thus we need to evaluate a single loop of Ω string. It is an easy calculation to evaluate a loop of Ω



(22.4)

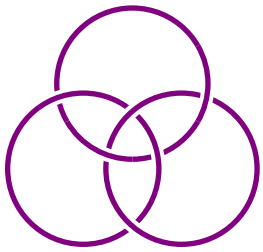


Fig. 22.11 Borromean Rings. Cutting any one strand disconnects the other two. Surgery on this link in S^3 creates the three-torus $S^1 \times S^1 \times S^1$.

Thus including the prefactor in Eq. 22.2 (the signature of the link is zero) we obtain

$$Z_{WRT}(S^2 \times S^1) = 1$$

which is in agreement with Eq. 7.8.

Finally let us consider the three-torus manifold $\mathcal{M} = S^1 \times S^1 \times S^1 = T^2 \times S^1 = T^3$. First, we note that surgery on the Borromean rings[21] (Fig. 22.11) yields the three torus (See exercise 22.3). To evaluate the link we use the corroloary of the killing property of the Ω strand,

[21] The rings are named for the crest of the royal Borromeo family of Italy, who rose to fame in the fourteenth century. However the knot (in the form of three linking triangles) was popular among Scandinavian runestones five hundred years earlier and were known as "Walknot" or "Valknut", or "the knot of the slain."

Fig. 17.8 to show (See exercise 22.4.)

$$Z_{WRT}(T^3) = \text{number of particle species} \tag{22.5}$$

which matches the prediction from Eq. 7.3 along with Eq. 7.5.

22.3.2 Turaev-Viro Revisited: Chain-Mail and the Turaev-Walker-Roberts Theorem

Using the ideas of surgery Roberts [1995] produced a beautiful geometric proof of the Turaev-Walker theorem (Turaev [1992, 1994]; Walker [1991]) which relates the Turaev-Viro invariant to the Chern-Simons (Witten-Reshitikhin-Turaev) invariant of a manifold. The result is, given a modular tensor category (a modular anyon theory) we have

$$Z_{TV}(\mathcal{M}) = |Z_{WRT}(\mathcal{M})|^2 \tag{22.6}$$

We will not give the full proof here, only the general idea.

First we will require one more minor collorary. Similar to Fig. 17.8 we have the identity shown in Fig. 22.12 (See exercise 22.6).

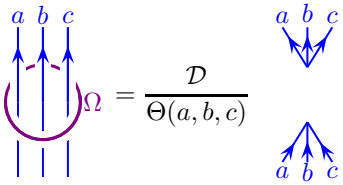


Fig. 22.12 The Ω strand fuses three lines due to the killing property. Here we have assumed an isotopy invarient theory as discussed in chapter 16, so we can draw vertices with all three lines pointing in the same direction so a, b, c fuse together to the identity.

We now want to construct a link of Ω strands which evaluates to the same value as the Turaev-Viro invariant discussed in chapter 21. Recall that to define the Turaev-Viro invariant, we first make a simplicial decomposition of the manifold, breaking it up into tetrahedra, we label each edge, and we sum a certain weight over all possible labelings as given in Eq. 21.1.

Given our simplicial decomposition here we will instead construct a link of Ω strands via the following procedure: Put one loop of Ω following the edges of each triangular face (colored gold in Fig. 22.13), and one loop of Ω around the waist of each edge (colored purple in Fig. 22.13) in such a way that the two types of strands link with each other. Such a link is known as *chain-mail*[22]. We then define the so-called *chain-mail invariant* of the manifold $\mathcal{M}$ as

$$CH(\mathcal{M}) = \mathcal{D}^{-N_v - N_{tet}} \left(\begin{array}{c} \text{Evaluate Chain-Mail Link of } \Omega \text{ strands} \\ \text{for simplicial decomposition of } \mathcal{M} \end{array} \right) \tag{22.7}$$

where N_v is the number of vertices in the simplicial decomposition and N_{tet} is the number of tetrahedra[23].

[22] When I have given talks on this subject I have been surprised to discover that many people don't know that chainmail is a medieval type of armor made of linked metal loops. Of course those who had misspent youth playing Dungeon's and Dragons, or reading the Lord of the Rings are very familiar with the concept and can tell you why Mithril is the best type of chainmail.

[23] More generally the chain-mail link can be defined for any *handlebody decomposition* of the manifold where Ω loops are put around 1-handles and 2-handles and N_v is then the 0-cells and N_{tet} is the 3-cells.

First, it is extremely easy to prove that the chain-mail invariant is independent of the particular simplicial decomposition (and hence is a manifold invariant as claimed). We need only show that it is unchanged under the Pachner moves (Fig. 21.4 and 21.5). This can be done entirely geometrically using only the killing property (Fig. 17.8) and the handle-slide property (Fig. 22.13) of the Ω strand (this is exercise 22.7).

Moreover, it is not hard to show that the chain-mail invariant is actually equal to the Turaev-Viro invariant. To do this we directly evaluate the chain-mail link. We start by using idenity 22.12 on each Ω strand attached to each face (those drawn as gold in Fig. 22.13). This generates a factor of $\mathcal{D}/\Theta(a,b,c)$ for each face. The remaining Ω strands (purple in Fig. 22.13) are decomposed into sums of all quantum numbers as per the definition of Ω in Fig. 17.6 each weighted by $d_a/\mathcal{D}$. This leaves one tetrahedron of particle strings per simplex as shown on the right of Fig. 22.13. (Note that the remaining tetrahedron of strings to be evaluated is a tetrahedral diagram *dual* to the original tetrahedron, in agreement with the discussion below Eq. 21.4).

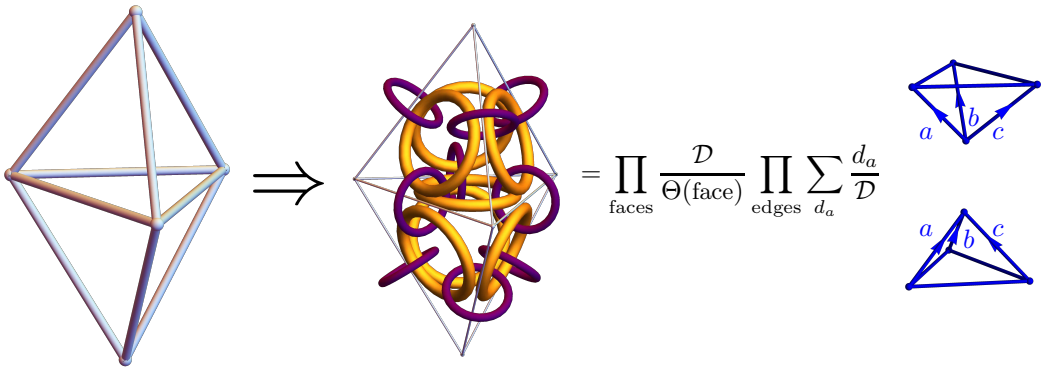


Fig. 22.13 The chain-mail invariant is equivalent to Turaev-Viro. We start with a simplicial decomposition on the left. To form the chain-mail link we put one Ω-loop around each triangular face (gold in the figure) and one Ω loop around each edge (purple in the figure) such that the gold and purple are linked. Let the gold loops "kill" the three purple strands that go through them using Fig. 22.12 to leave only tetrahedra (blue on the right) dual to the original tetrahedra.

Putting together the factors we have obtained leaves us with the chain-mail invariant (including the prefactor in the definition) being given by

$$CH(M) = \frac{\mathcal{D}^{N_f}}{\mathcal{D}^{N_v+N_e+N_{tet}}} \sum_{\text{edge labels}} \frac{\prod_{\text{tetrahedra}} \tilde{G}(\text{tetrahedron}) \prod_{\text{edges}} d_{\text{edge}}}{\prod_{\text{triangles}} \tilde{\Theta}(\text{triangle})}$$

with N_v, N_e, N_f, N_{tet} being the number of vertices, edges, faces (triangles), and tetrahedra respectively. Finally using the well-known topological fact that in three dimensions, the Euler characteristic $N_{tet} - N_f + N_e - N_v$ is zero, the factors of $\mathcal{D}$ are reassembeled to give exactly the

definition of the Turaev-Viro invariant Eq. 21.1 thus deriving

$$CH(M) = Z_{TV}(M)$$

Finally we turn to briefly discuss the derivation of the Turaev-Walker theorem Eq. 22.6. The key to this derivation is the fact[24] that if one uses a particularly simple decomposition of the manifold, surgery on the chain-mail link generates the connected sum of the original manifold $\mathcal{M}$ and its mirror image $\overline{\mathcal{M}}$

$$\text{Chain-mail link for } \mathcal{M} \overset{\text{surger}}{\rightarrow} \mathcal{M} \# \overline{\mathcal{M}} \tag{22.8}$$

[24] This key fact is not too hard to prove — it requires only about two paragraphs in the original workRoberts [1995]. However, it requires some knowledge of handlebody theory, so we will not discuss it here.

Evaluating the chain-mail link is therefore essentially equivalent to evaluating $Z_{WRT}(\mathcal{M} \# \overline{\mathcal{M}})$

Using the equivalence between chain-mail and the Turaev-Viro invariant we thus have (Eq. 22.3)

$$Z_{TV}(\mathcal{M}) \sim Z_{WRT}(\mathcal{M} \# \overline{\mathcal{M}}) \sim Z_{WRT}(\mathcal{M}) Z_{WRT}(\overline{\mathcal{M}}) \sim |Z_{WRT}(\mathcal{M})|^2$$

We have written this equation with $\sim$ rather than an equality because we have dropped factors of $\mathcal{D}$. To get these right we have to know more details about the particular decomposition of the manifold for which Eq. 22.8 holds so that we can keep track of the factors of $\mathcal{D}$ in the definition of the chain-mail invariant (Eq. 22.7). Keeping track of these factors carefully one obtains the desired Eq. 22.6.

Further Reading

Reshetikhin and Turaev [1991]; Lickorish [1993]; Witten [1989].

For more detailed discussion of Surgery and Kirby Calculus, as well as a nice discussion of manifold invariants, see Prasolov and Sossinsky [1996]

Roberts/Blanchet refinement.

Mention Crane-Yetter

The following references are standards for Surgery and Kirby Calculus, although they emphasize four dimensional topology. Gompf and Stipsicz [1999]; Kirby [1989]; Akbulut [2016].

Exercises

Exercise 22.1 Surgery on a Loop

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!).

Exercise 22.2 Surgery on the Hopf Link[Not hard if you think about it right!]

Consider two linked rings, known as the Hopf link (See Fig. 22.14). Consider starting with S^3 and embedding the 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.

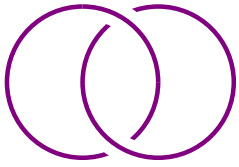


Fig. 22.14 A Hopf Link

Exercise 22.3 Surgery on the Borromean Rings[Hard]

(c) Consider the link shown in Fig. 22.11 known as the Borromean rings. 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 in the previous two problems. Show that one gets the three torus as a resul. Hint 1: Think about the group of topologically different loops through the manifold starting and ending at the same point, the so-called "fundamental group" or first homotopy group. (See section 33.3). 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 $aba^{-1}b^{-1}$. Hint 3: In some cases the fundamental group completely defines the manifold! (Don't try to prove this, just accept this as true in this particular case.)

Exercise 22.4 Evaluation of Borromean Ring Ω-Link

Use Fig. 17.8 to evaluate the Ω-link of Borromean rings shown in Fig. 22.11. Use this to establish Eq. 22.5. Note that the signature of the link is zero.

Exercise 22.5 Product of Blow Up and Blow Down

Use the handle-slide and the killing property of Ω to prove that the diagram made of two oppositely twisted Ω loops, as shown in Fig. 22.15, gives the identity.

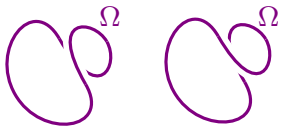


Fig. 22.15 The product of these two oppositely twisted Ω loops gives the identity.

Exercise 22.6 Killing Three Strands with Ω

Prove the relationship shown in Fig. 22.12.

Exercise 22.7 Pachner Moves and the Chain-Mail Invariant

Using killing moves (Fig. 17.8) and handle-slides (Fig. 22.15) show that the chain mail invariant Eq. 22.7 is unchanged under Pacher moves (Fig. 21.4 and 21.5). The answer is given by Roberts [1995], but it is a fun exercise. Looking up the answer spoils the fun!

Anyon Condensation

23

A physical mechanism that is very commonly discussed (in one language or another) is the idea of anyon condensation. The idea is modeled on the notion of conventional Bose-Einstein condensation. Under certain conditions one can imagine anyons forming a superfluid state, akin to a Bose-Einstein condensate. One can imagine making a condensate form either by continuously reducing the temperature with a fixed Hamiltonian, or by continuously changing the Hamiltonian at fixed (perhaps zero) temperature[1]. If one begins with a consistent anyon theory before the condensation, the system after the condensation will also be a consistent anyon theory, which we call the *condensed* theory[2]. It is believed that *all* continuous phase transitions that can occur between different anyon theories can be described in terms of anyon condensation[3].

There is a very detailed theory of anyon condensation, worked out by Bais and Slingerland [2009] and others (see references at end of chapter). Here we will give an abbreviated discussion, along with a few explicit examples.

Let us review some aspects of Bose condensation (See Leggett [2006] or Annette [2004] for much more information about the physics of superfluids and condensates). Recall that in a Bose condensate a macroscopic number of the particles reside in one special lowest-energy single-particle eigenstate which we call the condensate wavefunction. For a uniform system (say with periodic boundary conditions) the wavefunction for this single particle eigenstate is just a constant

$$\psi(\mathbf{r}) = \frac{1}{\sqrt{V}} \tag{23.1}$$

with V the volume of the system. It is crucial that bosons accumulate no phase or sign when they are braided around each other or exchanged with each other. If they were to accumulate any phase or sign, this would prevent them from remaining in the eigenstate Eq. 23.1 which is everywhere real and positive. This gives us:

Principle 1: Bosons must experience no net phase or sign when they move around then comes back to the same configuration — i.e., when bosons exchange or braid with other particles in the condensate.

Indeed, accumulating no sign when exchanging with other identical particles is the very definition of a boson[4].

With interacting bosons, one does not strictly have Bose condensation (not all of the bosons occupy the same single particle eigenstate, since

[1] A phase transition that occurs at zero temperature as some parameter of the Hamiltonian is changed is often known as a "quantum phase transition".

[2] It is sometimes possible that the condensed theory is a trivial theory — having only the vacuum particle type, and zero central charge. We should think of that as just being an uninteresting insulator. Strictly speaking this is a TQFT, just a very trivial one.

[3] First order, or discontinuous, phase transitions can always occur between any two phases of matter.

[4] It is possible that the condensate wavefunction has a spatial structure such as $e^{i\phi(\mathbf{r})}$, which happens when there is, say, a vortex within the condensate. What is crucial is that when bosons move around within the condensate, when they get back to the same many-particle configuration the phase is the same as when they started.

interactions kick the particles out of this eigenstate). Nonetheless, interacting bosons can condense to form superfluids which share many of the properties of Bose condensates. In particular, one still has the idea of a condensate wavefunction (or *order parameter*), and in order to form a condensate, no phase or sign must be accumulated when the particles exchange and braid.

To describe a condensate wavefunction (or order parameter) microscopically, one writes[5]

$$\phi(\mathbf{r}) = \langle \hat{\psi}(\mathbf{r}) \rangle$$

where $\hat{\psi}(\mathbf{r})$ is the (second quantized) operator which annihilates a particle at position $\mathbf{r}$. For noninteracting bosons, where many bosons are in the single particle wavefunction Eq 23.1, we obtain $|\phi|^2 = N_0/V$ where N_0 is the number of bosons in the single eigenstate.

The fact that this order parameter is (at least locally) number nonconserving (it destroys a particle) gives us the second important principle

Principle 2: Bosons can be freely absorbed by, or emitted from, the condensate.

[5]For systems with strictly fixed total number of particles this expectation would be zero and one instead looks at $\langle \hat{\psi}^\dagger(\mathbf{r})\hat{\psi}(\mathbf{r'})\rangle$ in the limit of $\mathbf{r}$ very far from $\mathbf{r'}$. This is known as "Off-Diagonal Long Ranged Order" or ODLRO.

23.1 Condensing Simple Current Bosons

We now would like to generalize the idea of Bose condensation to anyon theories. For simplicity we are going to restrict our attention to bosons that are also *simple currents*.[6] To remind the reader, a particle, let us call it J, is a simple current if $N_{Ja}^c = 0$ or 1 for all particle types a and c. This condition is equivalent[7] to the statement that $J^N = I$ for some integer N where I is the identity (where J^N here means N factors of J fused together).

For a particle J to condense, it must be a boson. This means that it must have trivial braiding with itself.

$$R_{J\times J}^{J,J} = 1$$

or equivalently a trivial spin factor (See Eq. 15.1)

$$\theta_J = 1$$

which is what we expect for a boson. This condition implements the above **Principle 1**: the boson must not experience a nontrivial phase as it exchanges with another particle as this would prevent a condensate wavefunction from forming.

Within the condensate, bosons may fuse with each other to form particles J^p for any value of p. It is not hard to show that all such resulting particle types must also be bosons $\theta_{J^p} = 1$ and further, they all must braid trivially with each other (See exercise 23.1).

While one *can* condense bosons that are not simple currents (See references at the end of the chapter), the rules for doing so are a bit more complicated and we will not discuss it here, as most of the important

[6]In the language of of conformal field theory, condensation of a simple current is known as "extension of the chiral algebra."

[7]This condition is also equivalent to either of the following also equivalent statements:

(a) $J \times \bar{J} = I$.

(b) $\mathsf{d}_J = 1$

physics is elucidated just for this special case.

We will start with an initial anyon theory which we call $\mathcal{A}$. Within this anyon theory, we assume there is a bosonic simple current J which we intend to condense to form a new anyon theory. The final anyon theory that comes out at the end of the condensation procedure will be called $\mathcal{U}$.

We can think of anyon condensation as proceeding in two conceptual steps as shown in Fig. 23.1. In between the initial theory $\mathcal{A}$ and the final theory $\mathcal{U}$ there is another theory $\mathcal{T}$ which is not a full anyon theory, but rather a fusion algebra (or planar diagram algebra), as we will discuss further in a moment.

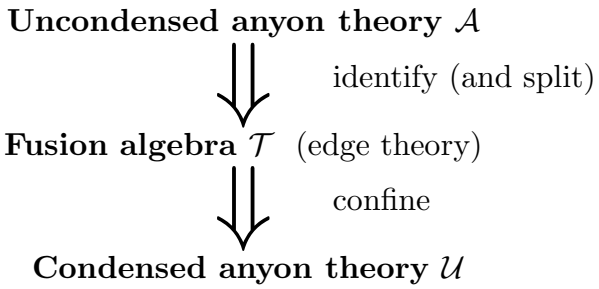


Fig. 23.1 Condensing one anyon theory to another can be described as having two "steps". The original anyon theory is labeled $\mathcal{A}$ and the final anyon theory is labeled $\mathcal{U}$. In between we have the intermediate theory $\mathcal{T}$ which is not generally a full-fledged anyon theory, but rather a fusion algebra (planar diagram algebra). The first step from $\mathcal{A}$ to $\mathcal{T}$ involves identification and possibly splitting. The second step from $\mathcal{T}$ to $\mathcal{U}$ involves confinement.

23.2 Identification Step

The first step in the condensation process is the identification step. In this step we group the particle types from the uncondensed theory $\mathcal{A}$ into so-called orbits.

Definition: The *orbit* of a particle type a under the action of J is the set of all particle types $b \in \mathcal{A}$ such that $b = J^p \times a$ for some integer p. We denote the orbit as $[a]^J$, or when it not ambiguous we just write $[a]$.

One should be cautioned[8] that the orbit $[a]$ is the same as orbit $[b]$ if $b = J^q \times a$ for any q.

Further, we note that if N is the smallest integer such that $J^N = I$ then there are at most N particle types in any given orbit, although there may be fewer particles in an orbit, as we will discuss in detail in section 23.4.

The physical point here is that all of the particles types in the same orbit of the original theory $\mathcal{A}$ are *identified* as being the same particle type in the $\mathcal{T}$ theory. The physical reason for this is **Principle 2**: bosons can be freely emitted from or absorbed into the condensate. A particular

[8] This can sometimes cause some notational confusion. It is often useful to choose a single representative of each orbit so that each orbit is uniquely denoted as a particular $[a]$ and one never writes $[b]$ if $b = J^q \times a$.

particle a can absorb a boson from the condensate and become $J \times a$ or it can absorb two bosons from the condensate to become $J^2 \times a$ and so forth. The quantum number a is no longer a conserved quantum number (and therefore is not a valid particle type), but the orbit $[a]$ remains conserved and can play the role of a particle type in the condensed theory. The orbits in the condensed theory will inherit fusion rules from the fusion rules of the uncondensed theory (with some potential complications we will address in section 23.4).

23.2.1 Orbits of maxmum size

Here we will consider the case where all of the orbits are of maximum size. I.e., all orbits have exactly N particle types in them (where N is the smallest positive integer so that $J^N = I$). We will return to the more complicated case where not all orbits are of maximum size in section 23.4.

We start with the original theory $\mathcal{A}$, and each particle a can be mapped to an orbit $[a]$ in the $\mathcal{T}$ theory. The fusion rules of the $\mathcal{T}$ theory are inherited from the fusion rules of the orginal anyon theory in a natural way which we can write in terms of the fusion multiplicity matrices as

$$N_{[a],[b]}^{[c]} = N_{a,b}^{c}$$

Note in particular that the identity particle I of the $\mathcal{A}$ theory maps to the orbit $[I]$ which becomes the identity particle of the $\mathcal{T}$ theory.

Example: $\mathbb{Z}_8^{(3+1/2)}$

Let us consider the anyon theory $\mathbb{Z}_8^{(3+1/2)}$ discussed in section 20.4.2 which is equivalent to the Chern-Simons theory $SU(8)_1$. There are 8 particles which we labelp $p = 0, \ldots, 7$ with fusion rules

$$p \times p' = (p + p') \text{ mod } 8 \tag{23.2}$$

and $p = 0$ is the identity. The corresponding twist factors are

$$\theta_p = \exp\left[\frac{2\pi i 7}{16} p^2\right] \tag{23.3}$$

We notice that $p = 4$ has trivial twist factor $\theta_4 = 1$ and is therefore a boson. Let us call this bosonic particle J, and we notice that $J^2 = I$ so the maximum orbit size is 2.

In this model we have four different orbits under the action of fusing with the boson J, and each of these orbits is of maximum size 2. Let us write down these orbits (Recalling that $[a]$ means the orbit of a)

[0]	which is also equal to [4]
[1]	which is also equal to [5]
[2]	which is also equal to [6]

[3] which is also equal to [7]

The meaning here should be obvious. Remembering that the boson J, which is particle $p = 4$ can be absorbed or emitted for free, we then for example, must identify particles 1 and 5 into a single orbit since fusing 1 with 4 gives 5 and fusing 5 with 4 gives 1.

These four different orbits comprise the particle types of the intermediate $\mathcal{T}$ theory. Let us denote these four orbits as $[p]$ with $p = 0, \ldots 3$. The fusion rules are inherited from the original uncondensed anyon fusion rules (Eq. 23.2) in an obvious way giving

$$[p] \times [p'] = [(p + p') \bmod 4] \tag{23.4}$$

with [0] playing the role of the identity in the $\mathcal{T}$ theory. To see how these fusion rules come from those of Eq. 23.2, consider, for example, $[1] \times [2] = [3]$: Here either 1 or 5 (the two particle types in the orbit [1]) fused with either 2 or 6 (the two particle types in the orbit [2]) will always give us 3 or 7 (the two particle types in the orbit [3]).

23.3 Confinement Step

The particle types in the intermediate theory $\mathcal{T}$ form a consistent fusion algebra (and indeed a consistent planar diagrammatic algebra) but they do not generally form a consistent anyon theory, as they do not have generally have a consistent braiding (or solution to the hexagon equations). The reason for this is that some of the particles in $\mathcal{T}$ are not valid particles of the final condensed anyon theory $\mathcal{U}$ and must be thrown out.

The reason some particles of $\mathcal{T}$ are not valid anyons in the condensed phase is that they braid nontrivially with the condensed boson. Trying to put a particle within the condensate that braids nontrivially with the condensed boson would violate **Principle 1**: when a boson in the condensate moves around, the phase must be the same when it arrives back at the same point. We thus have the rule that any particles a (or its orbit $[a]_J$) allowed in the final condensed theory $\mathcal{U}$ must braid trivially with the condensate, meaning that

$$R^{a,J}_{J\times a} R^{J,a}_{J\times a} = \frac{\theta_{J\times a}}{\theta_a} = 1 \tag{23.5}$$

where we have used Eq. 15.3 and the fact that $\theta_J = 1$. Since $J \times a$ and a are in the same orbit $[a]$ the condition Eq. 23.5 can be rephrased by saying that an orbit (a particle type of the $\mathcal{T}$ theory) is allowed into the final anyon theory $\mathcal{U}$ if all of the particles in the orbit have the same spin factor θ. Such particles that are allowed in $\mathcal{U}$ we say are *deconfined*, meaning that they can travel freely within the condensate. The particle types from $\mathcal{T}$ that braid nontrivially with the condesate are not allowed within the condensate and we say they are *confined*.

Although the confined particles of the $\mathcal{T}$ theory are not part of the

final condensed anyon theory, they still have physical meaning. The full $\mathcal{T}$ theory can be physically realized as a 1+1 dimensional theory living on the edge of a droplet of the $\mathcal{U}$ anyon theory living inside a larger region of the $\mathcal{A}$ uncondensed theory. The reason for this is that if we restrict motion of particles to a one-dimensional edge, there is no possibility of braiding one particle around each other and there is thus no problem with any of the particles of the full $\mathcal{T}$ theory — both the confined and deconfined particles can live there. The $\mathcal{T}$ theory, since it describes a 1+1 dimensional edge is not a braided anyon theory, but is rather a fusion algebra (or a planar diagram algebra).

If we try to drag one of the confined particles of the $\mathcal{T}$ theory into the condensed $\mathcal{U}$ droplet, its nontrivial braiding with the condensate creates a "branch-cut" in the condensate along its path into the condensate and destroys the condensate along this path. This costs an energy proportional to the distance the particle has been dragged into the $\mathcal{U}$ region. Thus there is a force pushing these confined particle back to the edge of the droplet. The particles are *confined* to the edge.

$\mathbb{Z}_8^{(3+1/2)}$ Again

Let us return to our example of $\mathbb{Z}_8^{(3+1/2)}$ and determine which of our orbits (particle types of $\mathcal{T}$) are confined or deconfined. Recall the rule that an orbit is deconfined if all of hte constituent particles in the orbit have the same twist factor θ. From Eq. 23.3 we have

$$\begin{aligned} \theta_0 &= \theta_4 = 1 \\ \theta_1 &\neq \theta_5 \\ \theta_2 &= \theta_6 = -i \\ \theta_3 &\neq \theta_7 \end{aligned}$$

Thus, the only two particle types allowed in the final condensed $\mathcal{U}$ theory are the orbits [0] (which is the identity) and [2] with corresponding spin factors $\theta_{[0]} = 1$ and $\theta_{[2]} = -i$. The only nontrivial fusion we obtain from Eq. 23.4 is

$$[2] \times [2] = [0].$$

We thus recognize this condensed anyon theory as the (left-handed) semion theory! Further we establish that if we condense a semion droplet within a $\mathbb{Z}_8^{(3+1/2)}$ background there will be two additional particle types ([1] and [3]) that remain confined at the edge of the droplet.

23.4 Splitting: Orbits not of maximum size

Above in section 23.2.1 we assumed all of the orbits were of maximum size. That is, if N is the smallest positive integer so that $J^N = I$, then all orbits are of size N.

If we have a situation where some orbits are not of maximum size, then we have a new physical phenomenon, known as *splitting*. This

phenomenon is a reflection of the fact that assigning each orbit $[a]$ from the uncondensed theory $\mathcal{A}$ to be a particle type of the intermediate theory $\mathcal{T}$ will not give an acceptable fusion algebra. Let us see how this happens.

Let us suppose we have some particle a such that $J^p \times a = a$ with $0 < p < N$ (In fact, p must divide N). We start by recalling

$$a \times \bar{a} = I + \dots$$

On the other hand, we can also write

$$a \times \bar{a} = (J^p \times a) \times \bar{a} = J^p \times (a \times \bar{a}) = J^p \times (I + \dots) = J^p + \dots$$

We thus conclude that we must have[9]

$$a \times \bar{a} = I + J^p + \dots \tag{23.6}$$

Now we claim that this feature will result in the orbits in $\mathcal{A}$ not producing an acceptable fusion algebra in $\mathcal{T}$. Let us see how this happen. As in the previous example, let us divide all of the particles in $\mathcal{A}$ into their orbits under the aciton of J, which we write as $[a]$. The fusion equation Eq. 23.6 then would imply the fusion for the orbits

$$\begin{aligned} [a] \times [\bar{a}] &= [I] + [J^p] + \dots \\ &= [I] + [I] + \dots \end{aligned} \tag{23.7}$$

where here we have used that $[J^p]$ is in the J orbit of the identity (i.e., $[J^p] = [I]$). Eq. 23.7 now presents an inconsistency as one of our rules of fusion algebras (Eq. 8.5) is that $N_{a\bar{a}}^I = 1$, i.e., the identity field should occur only once on the right hand side. We conclude that this is not acceptable as a fusion algebra for $\mathcal{T}$.

To resolve this problem the orbit $[a]$ must *split* into multiple particle types in $\mathcal{T}$ which we will write as $[a]_i$ with $i = 1, 2, \dots, q_a$ for some number q_a. Let us

Most generally we can write the mapping between the original $\mathcal{A}$ and $\mathcal{T}$ as

$$a \to \sum_{i=1}^{q_a} n_i^a \, [a]_i$$

where now $[a]_i$ are particle types of the $\mathcal{T}$ theory.

If the orbit of $[a]$ is maximal, then $[a]$ does not need to split, meaning ($n_1^a = 1$ and $q_a = 1$, and we don't need to write a subscript on $[a]$). However, if the orbit is not maximal, then $[a]$ must split into multiple different particles $[a]_1, [a]_2, \dots$ such that the twist factors all agree

$$\theta_{[a]_i} = \theta_a \quad . \tag{23.8}$$

As in the simple case with no splitting, the fusion rules of the $\mathcal{T}$ theory

[9] In fact we can generalize this argument to give

$$a \times \bar{a} = I + J^p + J^{2p} + \dots J^{N-p}$$

must be consistent with those of the uncondensed $\mathcal{A}$ theory. In particular

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

in the $\mathcal{A}$ theory implies

$$\left(\sum_{i=1}^{q_a} n_i^a \, [a]_i\right) \times \left(\sum_{i=1}^{q_b} n_i^b \, [b]_i\right) = \sum_c N_{a,b}^c \left(\sum_{i=1}^{q_c} n_i^c \, [c]_i\right)$$

within the $\mathcal{T}$ theory. This consistency implies the relationship

$$\mathsf{d}_a = \sum_{i=1}^{q_a} n_i^a \, \mathsf{d}_{[a]_i} \tag{23.9}$$

between the quantum dimensions in the $\mathcal{A}$ theory (left) and the quantum dimensions in the $\mathcal{T}$ theory (right). Once the particles have split, it is then possible to have a consistent set of fusion rules in the $\mathcal{T}$ theory. Once these fusion rules have been established, one can determine which fields are confined in order to determine the final condensed anyon theory $\mathcal{U}$.

While the phenomenon of identification (section 23.2) has a fairly obvious physical interpretation, it is often not as obvious how to interpret the phenomenon of splitting — except to say that it is required for consistency. However, a physical understanding is given by realizing the presence of a condensate can cause certain physical quantities to be locally conserved where they are indefinite in the uncondensed phase. It is the presence of these new locally conserved quantities which allow us to form $[a]_i$ where i can take q_a different values — corresponding to the q_a different values that the conserved quantity may take. This picture of emergent conserved quantities is elucitated by Burnell et al. [2011, 2012].

Example: $SU(2)_4$

Let us consider the example of the Chern-Simons theory $SU(2)_4$. We list the fields, their quantum dimensions, and their fusion rules in table 23.1.

We notice that particle 4 is a simple current with orbit of length 2 ($4 \times 4 = 0$). Let us now list the orbits in this theory

[0]	=	[4]	Maximum size orbit (also the identity)
[1]	=	[3]	Maximum size orbit
[2]			Not a maximum size orbit. This orbit must split

These orbits are just read off from the bottom line of the table 23.1: under fusion with the field 4, we have 0 mapping to 4 and vice versa; we have 1 mapping to 3 and vice versa, but 2 just maps to itself.

The fusion rules for the orbits, which we read off of table 23.1, are[10]

[10] Just to give an example, consider $[1] \times [1]$. Each factor of [1] could either represent the particle 1 or the particle $3 = 1 \times 4$. So the result of this fusion $[1] \times [1]$ could be $1 \times 1 = 0 + 2$ or $1 \times 3 = 2 + 4$ or $3 \times 1 = 2 + 4$ or $3 \times 3 = 0 + 2$. In all cases the result contains one particle from the [0] orbit and one from the [2] orbit. Thus giving $[1] \times [1] = [0] + [2]$.

particle	d	θ
0	1	1
1	$\sqrt{3}$	$e^{2\pi i/8}$
2	2	$e^{2\pi i/3}$
3	$\sqrt{3}$	$e^{2\pi i 5/8}$
4	1	1

$\times$	1	2	3	4
1	$0+2$	$1+3$	$2+4$	3
2	$1+3$	$0+2+4$	$1+3$	2
3	$2+4$	$1+3$	$0+2$	1
4	3	2	1	0

Table 23.1 Data for $SU(2)_4$. Left: quantum dimensions and twist factors for the different particles. Note that 0 is the identity. Right: Nontrivial fusion rules. Note that the fusion rules are given by Eq. 19.10, and the quantum dimensions are given by Eq. 19.2 given $d = \sqrt{3}$. You can check the consistency of d_a with the fusion rules by using Eq. 8.9 (i.e, d_a should be the largest eigenvalue of the fusion matrix $[N_a]^c_b$). Particle 4 is a simple currrent boson which we will attempt to consdense.

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

where [0] is the identity orbit. It is the last line here which demonstrates explicitly the problem noted in Eq. 23.7 — we should not have the identity twice on the right hand side. To fix this problem we split the particle [2] into two pieces $[2]_1$ and $[2]_2$.

$$[1]\times[1] = [0]+[2]_1+[2]_2 \quad (23.10)$$
$$[1]\times([2]_1+[2]_2) = [1]+[1] \quad (23.11)$$
$$([2]_1+[2]_2)\times([2]_1+[2]_2) = [0]+([2]_1+[2]_2)+[0] \quad (23.12)$$

While this is not quite the full fusion rules of the $\mathcal{T}$ theory, one can nonetheless extract[11] a unique set of fusion rules for the $\mathcal{T}$ theory consistent with Eq. 23.10-23.12, which are shown in table 23.2.

particle	d	θ
[0]	1	1
[1]	$\sqrt{3}$	$e^{2\pi i/8}$
$[2]_1$	1	$e^{2\pi i/3}$
$[2]_2$	1	$e^{2\pi i/3}$

$\times$	[1]	$[2]_1$	$[2]_2$
[1]	$[0]+[2]_1+[2]_2$	[1]	[1]
$[2]_1$	[1]	$[2]_2$	[0]
$[2]_2$	[1]	[0]	$[2]_1$

Table 23.2 Data for the intermediate $\mathcal{T}$ theory obtained from condensing the 4 particle in $SU(2)_4$. This is the fusion theory describing the edge of a condensed droplet.

[11] Eq. 23.10 is already written as a proper fusion rule. Eq. 23.11 implies $[1]\times[2]_1 = [1]$ and $[1]\times[2]_2 = [1]$. Next we note that the left hand side of Eq. 23.12 can be rewritten as $([2]_1\times[2]_1)+([2]_2\times[2]_2)+2([2]_1\times[2]_2)$ which, comparing to Eq. 23.12, immediately implies that $[2]_1\times[2]_2 = [0]$. To pin down the remaining fusion rule we use associativity $[2]_2 = ([2]_1\times[2]_2)\times[2]_2 = [2]_1(\times[2]_2\times[2]_2)$ which implies the only consistent set of fusion rules to be $[2]_2\times[2]_2 = [2]_1$ and $[2]_1\times[2]_1 = [2]_2$.

Several things are worth noting on this table. First, note that $d_{[2]_1} + d_{[2]_2}$ is the same as d_2 from the original $SU(2)_4$ as is required by Eq. 23.9. Secondly note that the twist factors are unchanged (even if a particle type splits) as stated in Eq. 23.8.

We can obtain the final $\mathcal{U}$ anyon theory, with a proper braiding, by throwing out the particles that are confined. Looking back at table 23.1 we see that the orbit [1] is made up of particles 1 and 3 which have different spin factors. This implies that particle [1] must be confined (it braids nontrivially with the condensate, see Eq. 23.5). However, the orbit [2] is made of a single particle type and therefore is deconfined (even if it splits). Thus our final anyon theory after confinement is given by table 23.3. We recognize the resulting anyon theory as $SU(3)_1$ or equivalently $\mathbb{Z}_3^1$. See section **??**.

particle	d	θ
$[0]$	1	1
$[2]_1$	1	$e^{2\pi i/3}$
$[2]_2$	1	$e^{2\pi i/3}$

$\times$	$[2]_1$	$[2]_2$
$[2]_1$	$[2]_2$	$[0]$
$[2]_2$	$[0]$	$[2]_1$

Table 23.3 Data for the final $\mathcal{U}$ anyon theory obtained from condensing the 4 particle from $SU(2)_4$. We recognize this theory as $SU(3)_1$ or equivalently $\mathbb{Z}_3^1$.

23.5 Other Features of Condensation

A few other features of condensation are worth mentioning. First, if we start with a modular anyon theory, $\mathcal{A}$, then the condensed theory $\mathcal{U}$ is also modular. Further for modular theories, the central charge (modulo 8) remains unchanged

$$c_{\mathcal{A}} = c_{\mathcal{U}} \mod 8 \tag{23.13}$$

Secondly, there is a beautiful relationship between the total quantum dimensions of the uncondensed theory $\mathcal{A}$, the fusion algebra $\mathcal{T}$ and the final theory $\mathcal{U}$. Recalling that total quantum dimension is defined by

$$\mathcal{D} = +\sqrt{\sum_a \mathsf{d}_a^2}$$

We then have

$$\frac{\mathcal{D}_{\mathcal{A}}}{\mathcal{D}_{\mathcal{T}}} = \frac{\mathcal{D}_{\mathcal{T}}}{\mathcal{D}_{\mathcal{U}}} \tag{23.14}$$

Let us check these relations for the $SU(2)_4$ condnensation example. From tables 23.1, 23.2, and 23.3 we obtain

$$\mathcal{D}_{\mathcal{A}=SU(2)_4} = \sqrt{12} \qquad \mathcal{D}_{\mathcal{T}} = \sqrt{6} \qquad \mathcal{D}_{\mathcal{U}=SU(3)_1} = \sqrt{3}$$

in agreement with Eq. 23.14. Also (From Eq. 17.14) we can calculate

that

$$c_{SU(2)_4} = c_{SU(3)_1} = 2 \pmod 8$$

in agreement with Eq. 23.13.

23.6 Cosets

In Chern-Simons theory, one of the most common ways to construct new TQFTs is the idea of a coset theory first discussed in the context of conformal field theory (See Di Francesco et al. [1997]; Moore and Seiberg [1989]; Goddard et al. [1985]). Given Lie groups G and H such that H is a subgroup of G, we may consider theories G_k (Chern-Simons theory G at level k) and correspondingly $H_{k'}$. There is then a well defined way to make a so-called *coset* theory which we write as $G_k/H_{k'}$. One rough physical interpretation of this construction is that we are *gauging* the subgroup H, essentially making these degrees of freedom redundant. One can have more complicated cosets where we embed H into a product of Lie groups $G \times G'$, to construct coset theories like $G_k \times G_{k'}/H_{k''}$.

While cosets of this type can seem quite complicated, they actually have an extremely simple interpretation in terms of boson condensation. To construct $G_k/H_{k'}$ we first construct $G_k \times \overline{H_{k'}}$ where the overbar means we should switch the chirality of the theory[12]. Then if we condense all possible simple current bosons in the product theory $G_k \times \overline{H_{k'}}$ we obtain the coset theory $G_k/H_{k'}$. (This is *much* simpler than the conventional coset construction!) This technique can be generalized in an obvious way. For example, to construct $G_k \times G_{k'}/H_{k''}$ we first construct $G_k \times G_{k'} \times \overline{H_{k''}}$ and then condense all of the simple current bosons. Note that given Eq. 23.13 and Eq. 17.23 the central charge of a coset is equal to the sum of central charges in the numerator minus the sum of central charges in the denominator.

[12] In Chern-Simons theory, $G_{-k} = \overline{G_k}$. Switching chirality in the diagrammatic algebra can be achieved by evaluating the mirror image of the diagram.

Example: $SU(2)_2/U(1)_2$

To construct the coset $SU(2)_2/U(1)_2$ we want to first construct $SU(2)_2 \times \overline{U(1)_2}$ and then condense all simple current bosons. Let us recall the data for $SU(2)_2$ (See section 18.3.1) and $\overline{U(1)_2}$ (See section 20.4.2) which are shown in Table 23.4.

The product theory $SU(2)_2 \times \overline{U(1)_2}$ has particles types of the form (a, b) where a is from the $SU(2)_2$ theory and b is from the $\overline{U(1)_2}$. The twist factor of such a product particle is $\theta_{(a,b)} = \theta_a \theta_b$.

One can see from the table that product particle $(\psi, 2)$ is a boson simple current (twist factor $\theta = 1$), so we can condense it. (In fact, not including the identity particle $(I, 0)$, this is the only boson simple current). The there are 12 particles in the product theory which divide into 6 orbits which all are of maximum size (so there is no splitting). These orbits (under the action of fusion with the $(\psi, 2)$ boson) are

$$[I, 0] = [\psi, 2] \quad ; \quad [\sigma, 0] = [\sigma, 2] \quad ; \quad [\psi, 0] = [I, 2]$$

$SU(2)_2$

particle	d	θ
I	1	1
σ	$\sqrt{2}$	$e^{2\pi i 3/16}$
ψ	1	-1

$\times$	σ	ψ
σ	I	σ
ψ	σ	$I+\psi$

$\overline{U(1)_2}$

particle	d	θ
0	1	1
1	1	$e^{2\pi i 7/8}$
2	1	-1
3	1	$e^{2\pi i 7/8}$

$i \times j = (i+j) \bmod 4$

Table 23.4 Data for $SU(2)_2$ (top) and $\overline{U}(1)_2$ (bottom). The overline indicates we take the mirror image theory, meaning all of the twist factors are complex conjugated compared to the definition given in section 20.4.2.

$$[I,1] = [\psi,3] \quad ; \quad [\sigma,1] = [\sigma,3] \quad ; \quad [\psi,1] = [I,3]$$

These are the six particle types of the $\mathcal{T}$ theory. Finally, examining the twist factors we can see that only the three orbits $[I,0]$ and $[\sigma,1]$ and $[\psi,0]$ are deconfined. The final condensed anyon theory $\mathcal{U} = SU(2)_2/U(1)_2$ is given in table 23.5. We recognize this result as being simply the Ising TQFT (See section 18.3.1).

particle	d	θ
$[I,0]$	1	1
$[\sigma,1]$	$\sqrt{2}$	$e^{2\pi i 1/16}$
$[\psi,0]$	1	-1

$\times$	$[\sigma,1]$	$[\psi,0]$
$[\sigma,1]$	$[I,0]$	$[\sigma,1]$
$[\psi,0]$	$[\sigma,1]$	$[I,0]+[\psi,0]$

Table 23.5 The final anyon theory data for the coset $SU(2)_2/U(1)_2$ is just the Ising TQFT.

Further Reading

Bais and Slingerland [2009] is the original discussion of anyon condensation. Full disclosure, I was the referee for this paper. I spent a lot of time reading it in great detail and I ended up deciding it was pretty brilliant.

A nice review of the physics of anyon condensation is given by Burnell [2018].

Eliëns et al. [2014] gives methods of extracting detailed data, such as F-matrices of a condensed theory from the data for the uncondensed theories.

Neupert et al. [2016] is also a good refernce on the mathematics of condensation.

Much of the structure of anyon condensation (at least for condensation of simple currents) was worked out originally in the context of so-called "fixed point resolutions" of conformal field theories. See for example, Schellekens [1999]

Chapter 32 shows how to use the computer program *Kac* to work out the results of certain condensations.

Exercises

Exercise 23.1 Fusion of Bosonic Simple Currents

Given an anyon theory with a bosonic simple current J, such that $R^{JJ}_{J^2} = 1$ and $J^N = I$, show that all of the particle types J^p with $0 < p < N$ are also bosons, and further that the braiding any two of these particle types is trivial $R^{J^p,J^{p+p'}}_{J^{p+p'}} = 1$. Hint: This is fairly trivial given some results of chapter 20.

Exercise 23.2 Splitting

Consider $SU(2)_2 \times \overline{SU(2)_2}$. There is a single simple current boson that can be consdensed. Find the $\mathcal{T}$ theory (there is a splitting!) and the final $\mathcal{A}$ theory after condensation.

Exercise 23.3 Cosets

(a) Calculate the properties of the coset $SU(16)_1/SU(2)_2$.

(b) Calculate the properties of the coset $SU(2)_1 \times SU(2)_1/SU(2)_2$.

Quantum Error Correction and The Toric Code

24

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.

24.1 Classical Versus Quantum Information

24.1.1 Memories

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

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$$

with arbitary complex prefactor α, β (where we normalize wavefunctions so $|\alpha|^2 + |\beta|^2 = 1$).

[2] Sometimes q-bit, but never cubit.

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

$$|\psi\rangle = \alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle \tag{24.1}$$

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!

24.1.2 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 24.1 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.

24.2 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).

24.2.1 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 corre-

[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?

spondingly the dimension of the Hilbert space is 2^N.

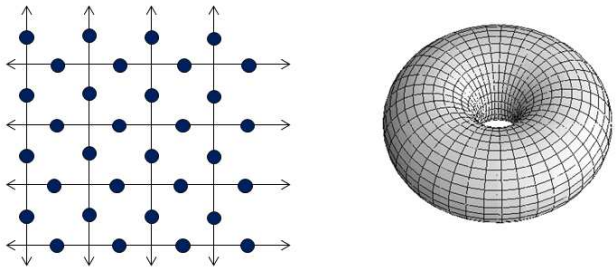


Fig. 24.1 The Hilbert space of the toric code — an N_x by N_y square lattice with spins (dots) on each edge wrapped up to make it periodic in both directions — i.e., a torus. Hence the name. There are 32 spins in this picture so the Hilbert space has dimension 2^{32}.

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. 24.2.

[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.

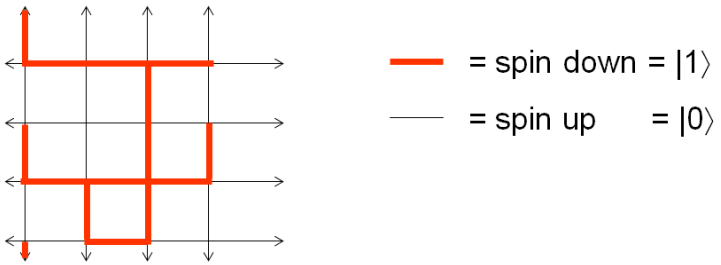


Fig. 24.2 A particular basis state of the Hilbert space, working in the up-dpwn basis (z-eigenstates). Here we denote down spins by thick (red) lines. And up spins are denoted by not coloring in the edges.

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.

24.2.2 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. 24.3. Note that there are a total of $N_x N_y$ vertex operators.

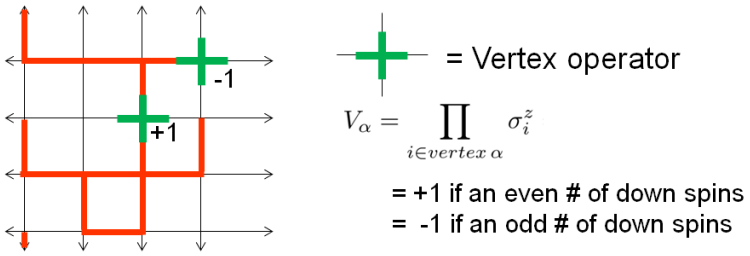


Fig. 24.3 The vertex operator returns $+1$ if there are an even number of incident down spins at that vertex and returns -1 if there are an odd number.

Note that it is possible (and useful) to define a corresponding projection operator

$$\tilde{V}_\alpha = \frac{1}{2}(1 - V_\alpha) \tag{24.2}$$

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. 24.4. There are a total of $N_x N_y$ plaquette operators.

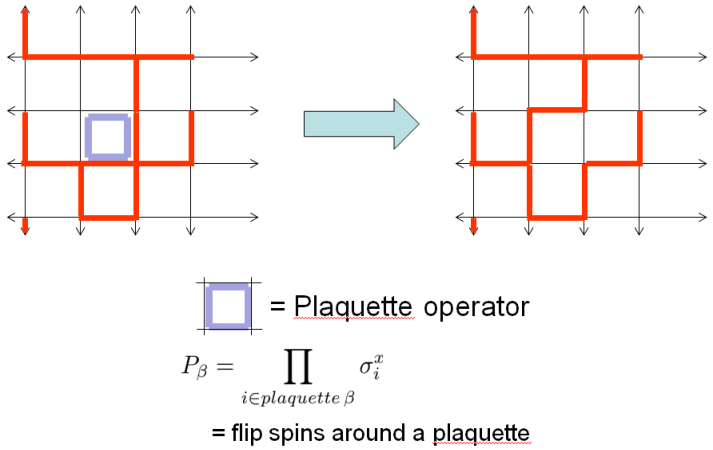


Fig. 24.4 The plaquette operator flips the state of the spin on the four edges of a plaquette.

As with the vertex operator, $P_\beta^2 = 1$ meaning P_β has eigenvalues $+1$ and -1. We can similarly define a projector

$$\tilde{P}_\beta = \frac{1}{2}(1 - P_\beta) \tag{24.3}$$

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. 24.5. The orthogonal superposition (adding the two states with a - sign) will give the other eigenstate.

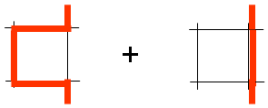


Fig. 24.5 A linear superposition of a flipped and unflipped plaquette is a $+1$ eigenstate of P_β or equivalently a 0 eigenstate of $\tilde{P}_\beta$. The -1 eigenstate is given by the orthogonal superposition, i..e, the superposition with a - sign between the two terms.

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 operators 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 a 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

$$\begin{aligned} &\text{Number of Vertex Ops} \quad + \quad \text{Number of Plaquette Ops} - 2 + 2g \\ &= \text{Number of Spins} \end{aligned}$$

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.

24.2.3 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.

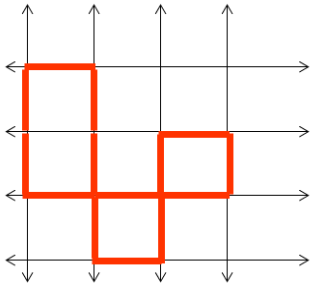


Fig. 24.6 A loop configuration consistent with the constraint that $V_\alpha = 1$ on every vertex. There must be an even number of red lines incident on every vertex.

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. 24.6

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. 24.5 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 configura-

tion. Then add to this in a superposition every configuration that can be obtained by flipping plaquettes. We thus have

$$|\psi\rangle = \sum_{\substack{\text{all loop configs that can}\\ \text{be obtained by flipping pla-}\\ \text{quettes from a reference}\\ \text{loop config}}} |\text{loop config}\rangle \qquad (24.4)$$

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. 24.7. If one flips a plaquette (blue in the fig) along this cut (green inn the fig), it does not change the parity of the number of red bonds that the cut goes through. Thus there are four independent wavefunctions of the form of Eq. 24.4, 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. 24.8

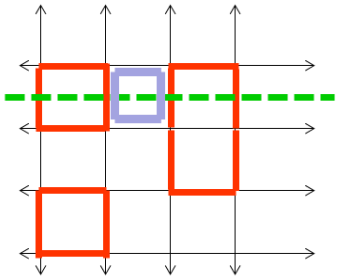


Fig. 24.7 Making a cut around one of the handles of torus, one can see that flipping a plaqutte, such as the blue one, does not change the parity of the number of red bonds cutting the green line. Further, it does not matter where (at which y-coordinate) the green cut is made, the number of red bonds it cuts is always even.

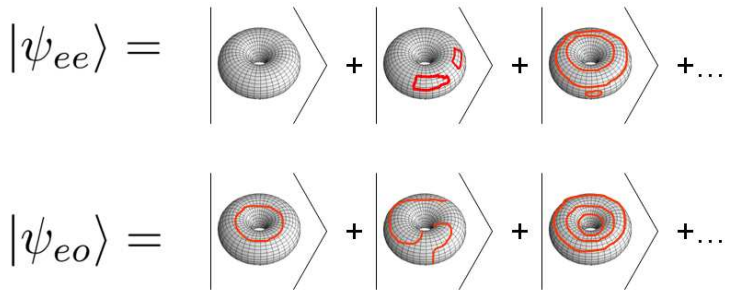


Fig. 24.8 Graphical depiction of $|\psi_{ee}\rangle$ which has an even number of strings running around each handle, and $|\psi_{eo}\rangle$ which is even around the first handle odd around the second.

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

$$|\psi\rangle = A_{ee}|\psi_{ee}\rangle + A_{eo}|\psi_{eo}\rangle + A_{oe}|\psi_{oe}\rangle + A_{oo}|\psi_{oo}\rangle \tag{24.5}$$

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. 24.1). We will refer to wavefunctions of the form of Eq. 24.5 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.

24.3 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.

24.3.1 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. 24.9.

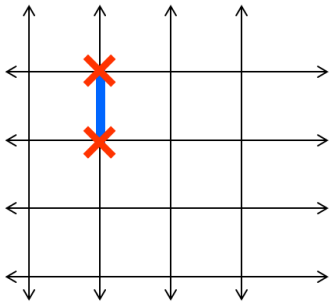


Fig. 24.9 A σ^x operator applied to the bond creates two vertices in the $V_\alpha = -1$ eigenstate.

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. 24.10, 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.

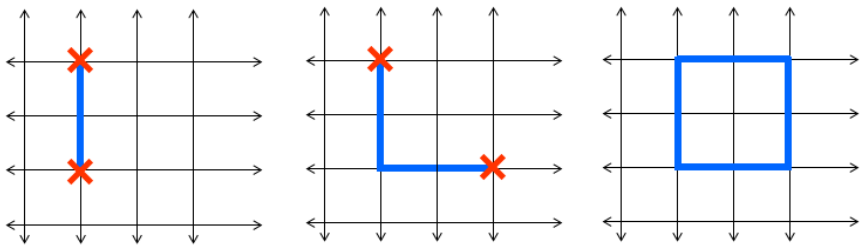


Fig. 24.10 Left: When two σ^x errors are made on bonds that share a vertex, the shared vertex is hit with σ_i^x twice, and thus becomes $V = +1$ again. Only the two vertices at the end of the "string" are in the $V = -1$ state. **Middle:** A longer string of errors. Note that we can only measure the 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 24.10 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. 24.10, 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. 24.10. 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 successfully 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 $V = -1$ vertices) 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.

24.3.2 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. 24.11. Recall that this eigenstate of the plaquette operator is a superposition of the flipped and unflipped plquettes similar to that shown in Fig. 24.5 but with a minus sign between the two terms.

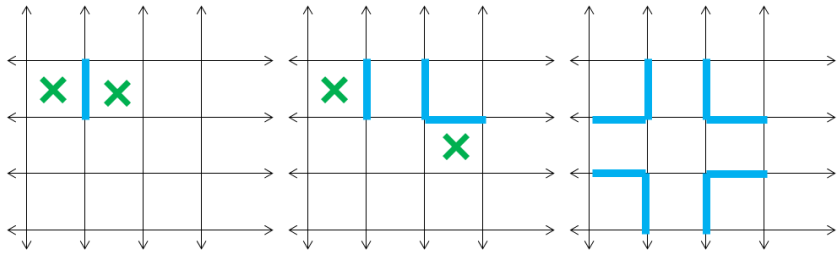


Fig. 24.11 Left: When a σ^z error is applied to a bond, the plaquettes on either side end up in the $P = -1$ state **Middle:** A string of several σ^z errors. **Right** A closed loop of σ^z errors. This is equal to the product of all of the enclosed V_α operators. In the code space, this is equal to $+1$.

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. 24.11. 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. 24.11, this loop

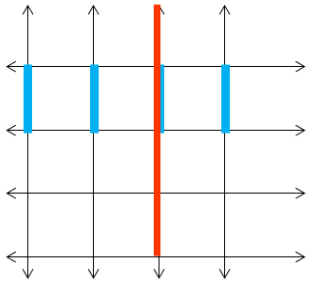


Fig. 24.12 If a string of σ^z goes around a handle, it measures the parity of the number of red strings going around the dual handle.

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. 24.12, this operator then counts the parity of the number of red bonds going around the dual handle, as shown in the figure. Thus, applying the string of σ^z operators around the handle makes the transformation

$$A_{ee}|\psi_{ee}\rangle + A_{eo}|\psi_{eo}\rangle + A_{oe}|\psi_{oe}\rangle + A_{oo}|\psi_{oo}\rangle \longrightarrow$$
$$A_{ee}|\psi_{ee}\rangle + A_{eo}|\psi_{eo}\rangle - A_{oe}|\psi_{oe}\rangle - A_{oo}|\psi_{oo}\rangle$$

24.3.3 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!

24.3.4 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 measurement of the logical bits.
(2) As mentioned above, the toric code as a 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 the 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. (CITE)
(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.

24.4 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

$$H = -\sum_{vertices\ \alpha} V_\alpha - \sum_{plaquettes\ \beta} P_\beta \tag{24.6}$$

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. 24.2 and 24.3. Writing

$$\tilde{H} = \sum_{vertices\ \alpha} \tilde{V}_\alpha + \sum_{plaquettes\ \beta} \tilde{P}_\beta \tag{24.7}$$

which differs from Eq. 24.6 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.

24.4.1 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 the table in Fig. **??**.

$\times$	I	e	m	f
I	I	e	m	f
e	e	I	f	m
m	m	f	I	e
f	f	m	e	I

Fig. 24.13 Fusion Table for the Toric Code

24.4.2 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].

[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

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 shown in the right panel of Fig. 24.10 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. 24.11 and is the product of a string of σ^z operators. Recall

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 15.2. Considering Fig. 24.14

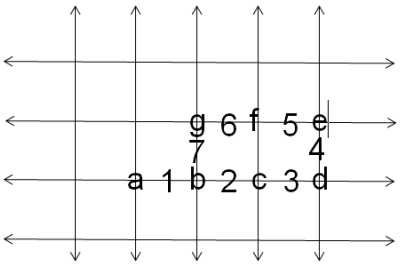


Fig. 24.14 Vertices are labeled with letters and bonds are labeled with numbers.

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 15.2. 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.

that the reason this process does not accumulate a phase is because the string of σ^z operators around the loop is equivalent to the product of the V_α vertex operators enclosed — and in the ground state, the V_α operators are in the $+1$ state. However, if there is one e particle inside the loop, this means that one of the V_α operators is actually in the -1 state. In this case the phase of taking the m particle around in a loop is actually -1. So there is a phase of -1 for taking m around e.

Properties of f, the fermion

Since f is made up of an m bound to an e, it is easy to see that taking e around f accumulates a phase of -1 and taking m around f also accumulates a phsae of -1. More interesting is the properties of a single f. We claim that f is a fermion. The easiest way to see this is to check its phase under a twist as shown in Fig. 24.15

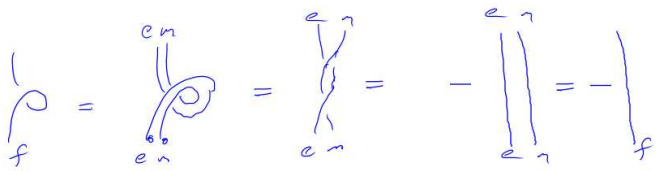


Fig. 24.15 The $f = e \times m$ particle is a fermion, since e braiding around m gives a -1 sign.

Note that taking f all the way around f will result in a net $+$ sign.

24.4.3 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 $\mathcal{D} = 2$.

24.4.4 Flux Binding Description

We can describe the physics of the toric code phase in a flux binding description somewhat analogous to Chern-Simons theory. Here let us

define

electric particle = e = particle bound to 1 unit of electric charge

magnetic particle = m = particle bound to π units of magnetic flux

fermion = f = particle bound to 1 unit of electric charge and π units of magnetic flux

It is easy to see that this charge and flux will correctly give the +1 and -1 phases accumulated from braiding particles.

24.5 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}

[6] This is a Brillouin-Wigner perturbation theory, where successive terms are rigorously λ/Δ smaller.

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 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 **??** — it is not possible that the braiding phases in a theory change an arbitrarily small amount.

24.6 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

25

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. 25.1

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 element $g \to g^{-1}$ as shown in Fig. 25.2.

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. 25.3. This is the string-net vertex fusion rule.[2]

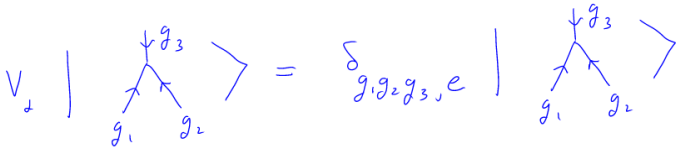


Fig. 25.3 Definition of V_α when all arrows are directed into the vertex (if a vertex is directed out, one can invert the arrow and invert the group element). The vertex operator gives zero unless the product of group elements around the plaquette gives the identity element e

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. 25.4.

[1] I present this model on the "dual" graph compared to Kitaev's presentation.

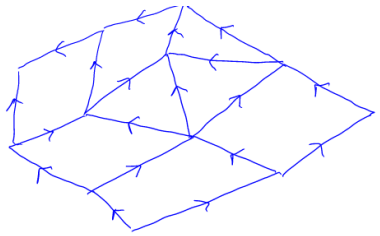


Fig. 25.1 Part of a directed graph.

Fig. 25.2 Inverting the direction on an edge inverts the group element.

[2] Note that if we have three edges coming into a vertex labeled g_1, g_2, g_3 the condition $g_1 g_2 g_3 = e$ is equivalent to $g_2 g_3 g_1 = e$ and $g_3 g_1 g_2 = e$.

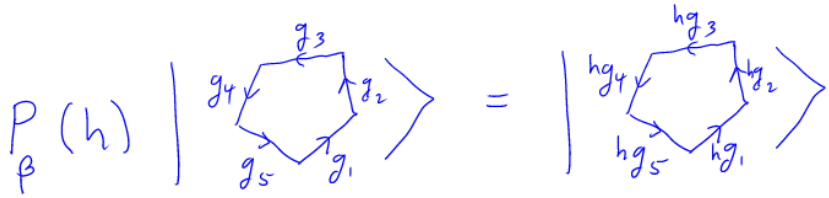


Fig. 25.4 The plaquette operator $P_\beta(h)$ premultiplies all of the clockwise oriented bonds by the element h.

The total plaquette operator (the one that will enter the Hamiltonian) is then defined to be

$$P_\beta = \sum_{g \in G} P_\beta(g)$$

It is easy to see that the plaquette operator and the vertex operator commute.

Relation to toric code

How does this related to the toric code? Consider the group $\mathbb{Z}_2$ of two elements where we write the two elements as $\{1, -1\}$. We can think of these as being spin up and spin down on the lattice. Since $g = g^{-1}$ for every element we don't need to put arrows on the lattice.

$$\begin{aligned} P_\beta(1) &= \text{identity operator} \\ P_\beta(-1) &= \text{multiply all edges by -1. (i.e. flip all edges)} \end{aligned}$$

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.

Working with abelian groups, this new toric code is a fairly straightforward generalization of the toric code we have already studied. However, 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[3]. Generically one will have nonabelian anyons. I will not go through this argument in detail. See Kitaev for more. (Also cite Propitius)

This model by Kitaev is essentially a lattice gauge theory. Essen-

[3]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 $ug = gu$.

tially the wavefunction is given by a unique state plus everything that is "gauge equivalent" (meaning can be obtained by plaquette flips). Let us think in terms of the dual lattice for a moment (so plaquettes become dual-vertices and vertices become dual-plaquettes). The sum over group elements of $P_\beta(h)$ enforces gauge invariance of the theory at the dual vertices. The vertex operator V_α then assures there is no magnetic flux penetrating the dual plaquette.

25.0.1 $\mathbb{Z}_N$ toric code

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.

Merge these

Perhaps the most simple generalization of the toric code is the $\mathbb{Z}_N$ toric code. Here each edge of the lattice is labeled with an element of the group $\mathbb{Z}_N$, i.e., an integer modulo N with the group operation of addition. It is easy to work out that one obtains a correpsonding $\mathbb{Z}_N$ electric charge at the vertices and $\mathbb{Z}_N$ magnetic charge on the plaquettes. Let us call the elementary electric charge e, and the elmentary magnetic charge m. These have the property that $e^N = m^N = I$. We then have N^2 particle types that we can label as

$$(i, j) = e^i \times m^j$$

with i and j being chosen from $0, \ldots (N-1)$, with the fusion rules corresponding to just addition of the i and j indices modulo N. The e and m particles again are bosons (they braid trivially with themselves). However, as in the $\mathbb{Z}_2$ case braiding an e around an m is nontrivial[4]: here it gives a phase of $e^{2\pi i/N}$.

[4] One might wonder where we have broken time reversal to get a complex phase. In fact we have not broken time reversal — however, we did have to make a choice as to which particle we would call e. If we had chosen the particle $e^{N-1} = e^{-1}$ instead to be the elementary particle, the phase would be reversed.

25.1 Ground State Degeneracy in the General Nonabelian Case

While the full particle spectrum for the Quantum Double of a nonabelian group is tricky to calculate, it is not too difficult to calculate the ground state degeneracy on a torus (and hence determine the number of anyons in the theory). Here we use to our advantage that we can use any lattice we like to cover the torus, so we might as well choose a simple one like that shown in Fig. 25.5. In that lattice covering of the torus, there are two vertices, and the vertex operator then requires that $abc = I$ and $a^{-1}b^{-1}c^{-1} = I$. This then implies that

$$ab = ba = c^{-1}$$

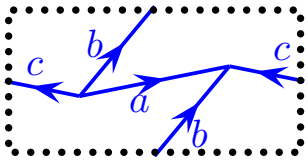


Fig. 25.5 The simplest possible decomposition of a torus into a single plaquette with three edges and two trivalent vertices. Here the dotted lines are periodic boundary conditions on the torus. The edges are labeled with particle types. The vertex conditions require $abc = I$ and $a^{-1}b^{-1}c^{-1} = I$.

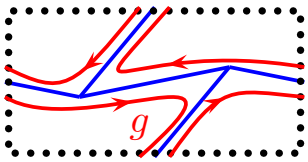


Fig. 25.6 The simplest possible decomposition of a torus into a single plaquette with three edges and two trivalent vertices. Here the dotted lines are periodic boundary conditions on the torus. The edges are labeled with particle types. The vertex conditions require $abc = I$ and $a^{-1}b^{-1}c^{-1} = I$.

and in particlular a and b must commute.

Now let us think about the plaquette operator. In this lattice three is only a single

Use Burnsides lemma (otherwsie known as the lemma not by Burnsides)

More Generalizing The Toric Code: Loop Gases and String Nets

26

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.

26.1 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. 26.1 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

$$|\psi\rangle = \sum_{\substack{\text{all loop configs that can be} \\ \text{obtained from a reference} \\ \text{loop config}}} |\text{loop config}\rangle \qquad (26.1)$$

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.

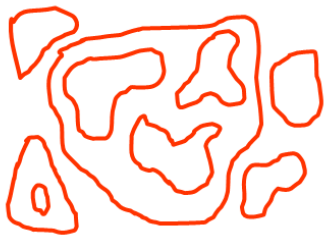


Fig. 26.1 A loop gas in 2d. We can think of this as particle world-lines in 1+1 d.

Move 1: "Isotopy" = smooth deformation of a loop. As shown in Fig. 26.2. We have always allowed smooth deformations in our diagrammmatic algebras.

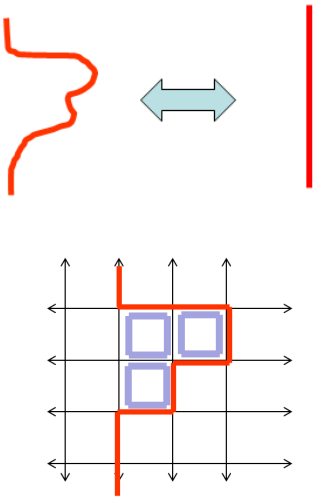


Fig. 26.2 Isotopy (Top) Off the lattice this is just deformation of a line. (Bottom) on the lattice, this is implemented by flipping over the blue plaquettes.

Move 2: "Adding or removing a loop". As shown in Fig. 26.3

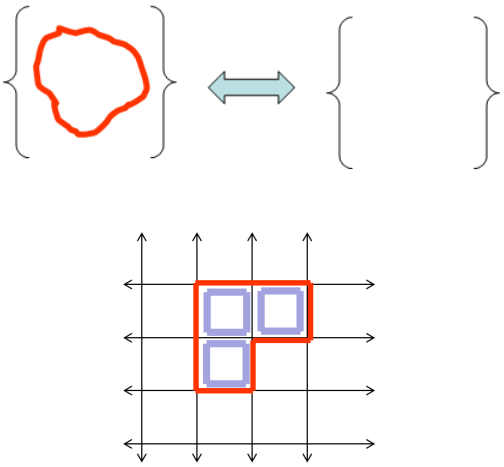


Fig. 26.3 Adding or Removing a loop (Top) Off the lattice (Bottom) On the lattice we flip the shown plaquettes.

Move 3: "Surgery" or reconnection of loops. As shown in Fig. 26.4

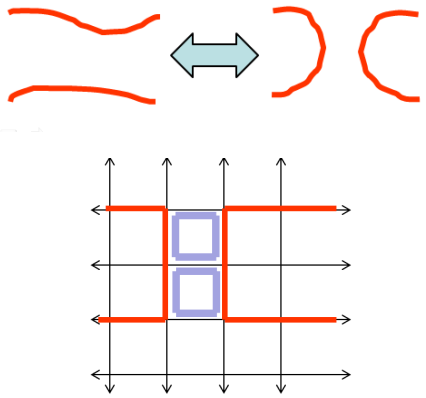


Fig. 26.4 Loop Surgery (Top) Off lattice surgery (Bottom) On lattice, flip the shown plaquettes

We can summarize these rules with simple skein-like relations as shown in Fig. 26.5

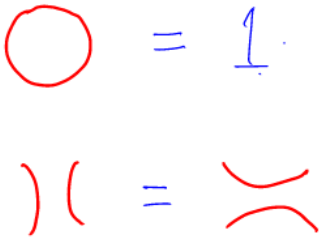


Fig. 26.5 "Skein" relations for the toric-code loop gas. The unity on the right of the top line means that the amplitude in the superposition that forms the wavefunciton is unchanged (multiplied by unity) under removal or addition of a loop.

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.

26.1.1 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. 26.6.

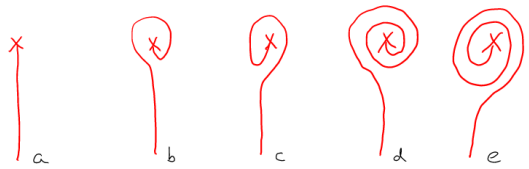


Fig. 26.6 Ends of strings can be wrapped either way, and multiple times. a and b are different, c is equivalent to b by surgery. Similarly d and e are both the equivalent to a.

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

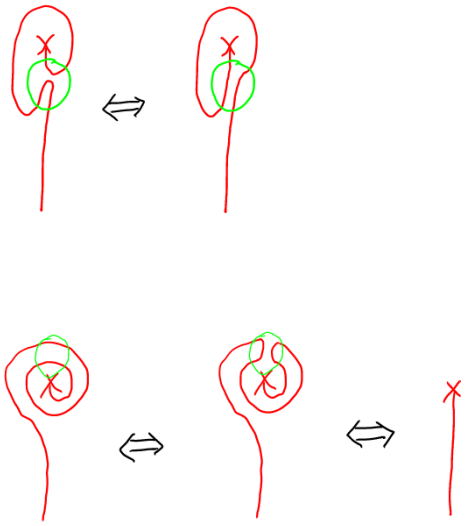


Fig. 26.7 Loop equivalences. Surgery is done inside the light green circles. The final equality on the lower right is just pulling the string tight.

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. 26.8

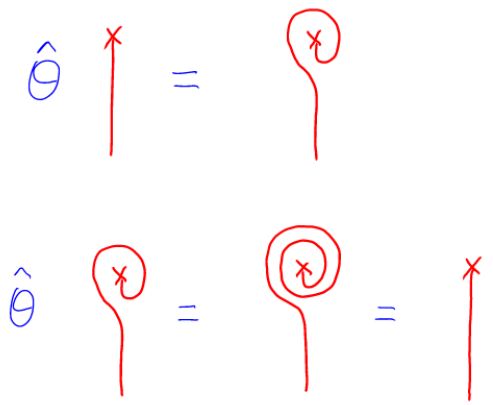


Fig. 26.8 Rotation

From these relations we can determine that the eigenvalues of the rotation operator are $+1$, correpsonding to the e particle and -1 corresponing to the f particle, as shown in Fig. 26.9.

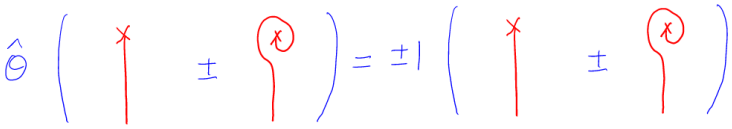


Fig. 26.9 The eigenvectors of the rotation operator $\hat{\theta}$

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. 26.10.

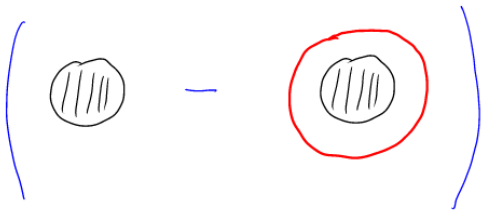


Fig. 26.10 The black disk is some region of our model. Forming a superposition of this region, and this region with a loop around it, with a minus sign between the two pieces, must be orthogonal to the ground state — it puts a magnetic excitation m in the region.

26.2 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.

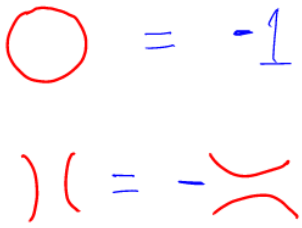


Fig. 26.11 "Skein" relations for the double-semion loop gas. Each loop removal/addition and each surgery incurs a minus sign. Note that these are the same as the Kauffman rules when we considered semions.

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

$$|\psi\rangle = \sum_{\substack{\text{all loop configs that can be} \\ \text{obtained from a reference} \\ \text{loop config}}} (-1)^{\text{Number of Loops}} |\text{loop config}\rangle \tag{26.2}$$

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.

26.2.1 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.26.12). 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. 26.12. The advantage of this structure is that loops cannot intersect as they can (at the 4-fold corner) on the the square lattice.

Fig. 26.12 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. 26.13. Depending on the initial state, when the plaquette is flipped, one may or may not obtain a minus sign.

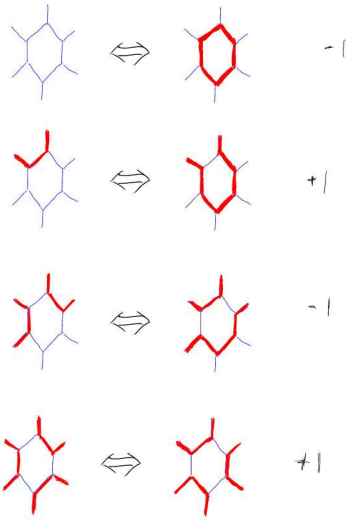


Fig. 26.13 Some plaquette flips for the double semion model on the hexagon. The top line obviously adds a loop, so should get a minus sign. The second line just stretches a loop over a plaquette, so does not get a minus sign. The third line is a surgery so gets a minus sign. The fourth line is a double 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.

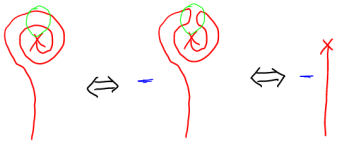


Fig. 26.14 Surgery incurs a minus sign. Compare to fig. 26.7

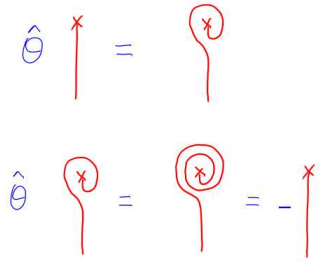


Fig. 26.15 Surgery incurs a minus sign. Compare to fig. 26.7

26.2.2 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. 26.14 Resulting in the effect of rotation being Fig. 26.15 Again we can use these to give us the eigenstates of the rotation operator as shown in Fig. 26.16

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!

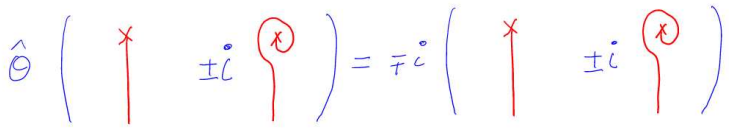


Fig. 26.16 Eigenstates of the rotation operator for the doubles semion model.

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. 26.10, 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 given in Fig. 26.17.

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. 26.10 (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.

$\times$	I	ϕ	ϕ^*	m
I	I	ϕ	ϕ^*	m
ϕ	ϕ	I	m	ϕ^*
ϕ^*	ϕ^*	m	I	ϕ
m	m	ϕ^*	ϕ	I

Fig. 26.17 Fusion Table for Double Semion Model

[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 **??**, and gives $\mathcal{D}$ on the vacuum.

26.3 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

Fig. 26.18 A general string net, that allows branching, here with two colors.

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. 26.18. 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!).

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. 26.19.

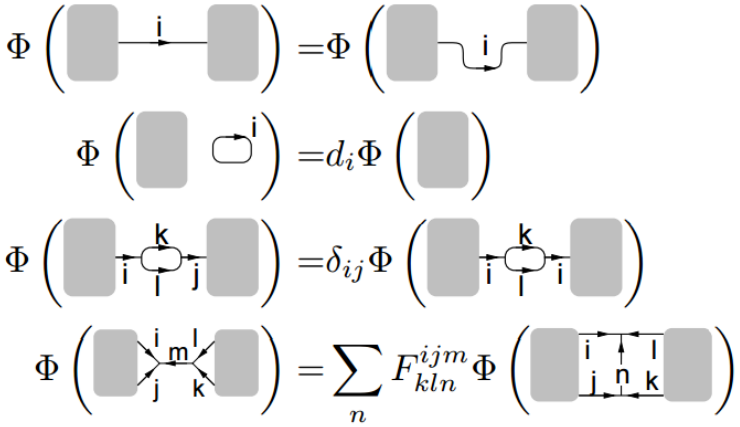


Fig. 26.19 Rules for a string net. The grey regions are meant to be the same on both the left and the right of the diagram. Figure stolen from Levin and Wen.

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. 26.5 and Fig. 26.11.

It is important to note that the F-matrix used to define define the string net (last line of Fig. 26.19) 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.

26.4 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. 26.20.

Fig. 26.20 A branching string net for the doubled Fibonacci model.

Starting with Eq. 9.2, we consider the following F-moves as shown in Fig. 26.21

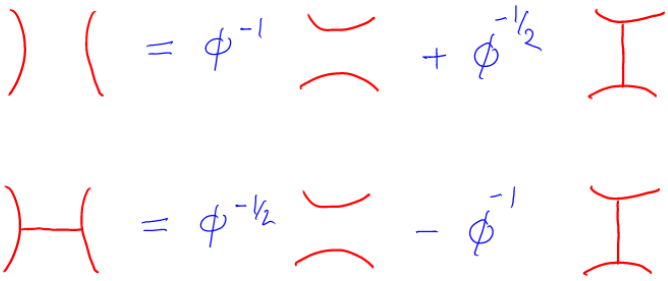


Fig. 26.21 Rules for building the doubled Fibonacci model.

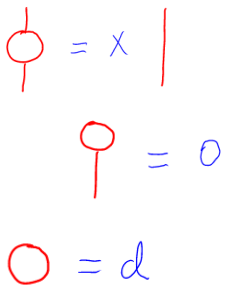


Fig. 26.22 Rules for building the doubled Fibonacci model.

Where here $\phi = (1+\sqrt{5})/2$ and (the values of these coefficients come from the values of the F-matrix in Eq. 9.2.

We also expect to have rules of the form of Fig. 26.22 The first and second rules[2] are results of locality. The final rule is the usual rule that a

[2] In fact we can prove that the tadpole rule must be zero. This is a homework problem!

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. 26.21 to give Fig. 26.23.

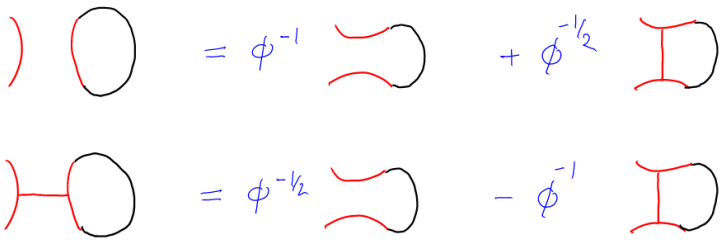


Fig. 26.23 Starting with Fig. 26.21 and closing strings to the right hand. The black strings should be imagined to be red — they are drawn black so one can see what is added compared to Fig. 26.21

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. 26.21 and 26.22. We can then write a ground state wavefunction of the form

$$|\psi\rangle = \sum_{\substack{\text{all string net configs that} \\ \text{can be obtained from a ref-} \\ \text{erence config}}} \Phi(\text{net config}) \ |\text{net config}\rangle$$

This looks quite similar to our above toric code loop gas, except now we allow branching string nets instead of just loops, and also the kets have a prefactor Φ. These prefactors are chosen such that the algebraic rules described above are satisfied. I.e., removing a loop increases Φ by a factor of d. Removing a bubble (as in the upper left of 26.22) 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. 26.21.

26.4.1 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. 26.24 Just as an example, consider the ending shown on the left of Fig. 26.25. By using an F-move, it is reduced to a combination of the three presented above.

Fig. 26.24 Possible string endings in the doubled Fibonacci string net model.

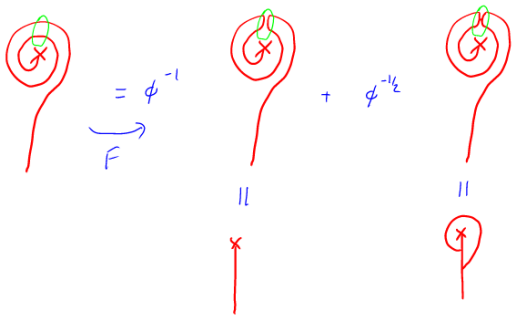


Fig. 26.25 An example of reducing a more complicated string ending into one fo the three endings shown in Fig. 26.24.

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. 26.26 and then using F-matrices to reduce the result back to linear combinations of the same three possible endings.

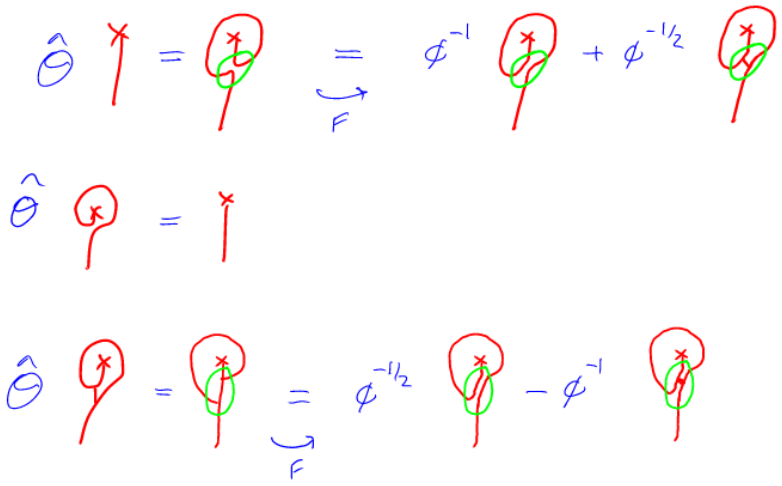


Fig. 26.26 The rotation operator $\hat{\Theta}$ applied to the possible string endings. Then using F matrices we reduce the results to linear combinations of the same endings.

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 13.3.). 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!).

26.4.2 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. 26.27

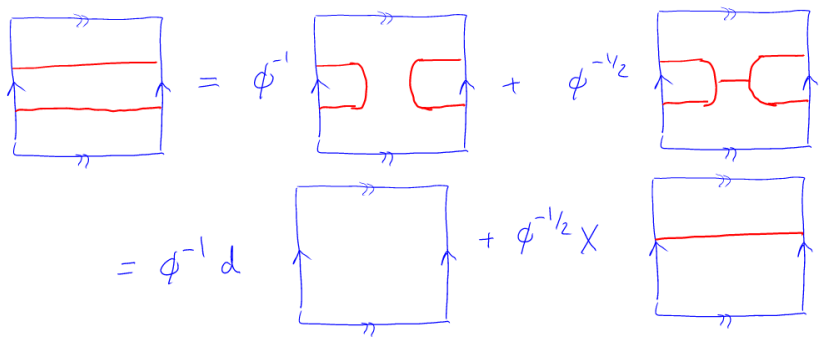


Fig. 26.27 Reducing two loops around a handle to a linear combination of one loops and zero loops.

26.5 Add details of Levin Wen Model on the Lattice?

26.6 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. 26.28

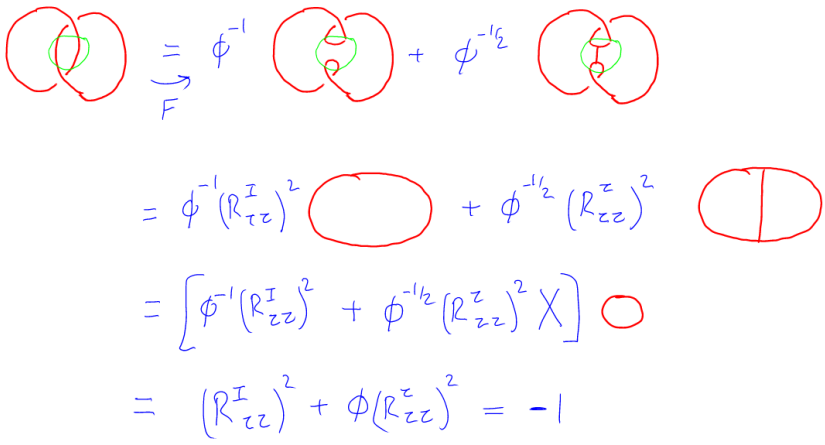


Fig. 26.28 Calculating the nontrivial element of the Fibonacci anyon S-matrix.

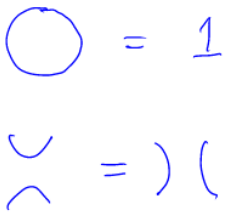


Fig. 26.29 Loop gas rules for the toric code

Exercises

Exercise 26.1 *Quasiparticles in Toric Code Loop Gas*

As discussed in lecture, the toric code ground state can be considered to be a loop gas with the rules given in Fig. 26.29

Certain quasiparticle excitations can be indicated as ends of strings in the loop gas.

(a) Show that the linear combinations of string ends shown in the figure 26.30 are eigenstates of the rotation operator – with the boson accumulating no phase under rotation and the fermion accumulating a minus sign. (We did this in lecture so it should be easy).

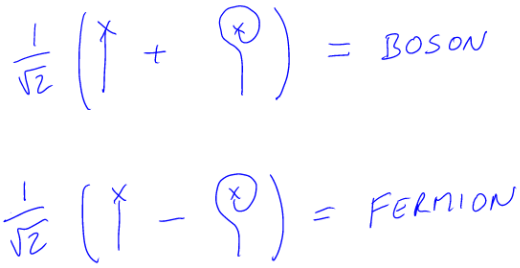


Fig. 26.30 Boson and Fermion quasiparticles as string ends in the toric code loop gas

(b) Consider exchanging two such quasiparticles. To get a general idea of how the calculation goes, you will have to evaluate diagrams of the form of

Fig. 26.31. Show that one obtains bosonic or fermionic exchange statistics respectively for the two linear combinations shown above.

(c) [Harder] Consider fusing the boson (the electric particle e) and the fermion together. Show that this creates a magnetic defect which does not have a trailing string. You will have to recall that the operator that creates a magnetic particle is sum of the identity operator and minus an operator that draws a loop all the way around the region. (This operator is a projector that forces a magnetic defect into a region; the orthogonal projector assures that there is no magnetic defect within the region).

Fig. 26.31 Braiding defects

Exercise 26.2 *Quasiparticles in Double Semion Loop Gas*

As discussed in lecture, the doubled semion model ground state can be considered to be a loop gas with the rules given in Fig. 26.32. Note that these rules are the same as the semion rules from the problem "Abelian Kauffman Anyons" which we considered earlier (although in that model there is only one chirality of semion particle!)

Again certain quasiparticle excitaitons can be indicated as ends of strings in the loop gas.

(a) Show that the linear combinations of string ends shown in the figure 26.33 are eigenstates of the rotation operator – with the two particles accumulating a factor of i or $-i$ under rotation (We also did this in lecture so it should be easy).

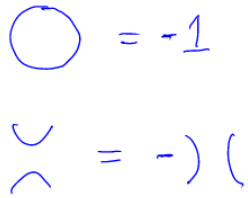


Fig. 26.32 Loop gas rules for the doubled semion model

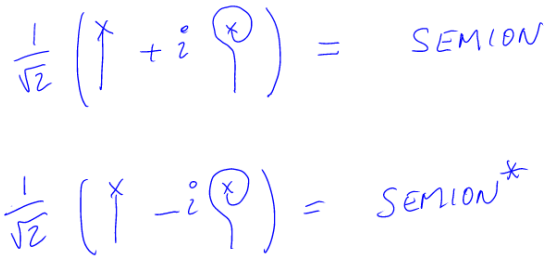


Fig. 26.33 Semion and anti-semion string ends in the doubled semion loop gas

(b) Consider exchanging two such quasiparticles. Show that under exchange one obtains factor of i or $-i$ as expected for semions and anti-semions. Note: The anti-semion is not the antiparticle of the semion (I know it is bad nomenclature!) – The antisemion is the opposite handed particle. The semion is its own antiparticle.

(c) [Harder] Consider fusing the semion and anti-semion together. Show that this creates a "magnetic defect." What is the projector that produces a magnetic defect in a region?

Exercise 26.3 *Double Fibonacci String Net*

(a) As discussed in lecture, the double Fibonacci model ground state can be viewed as a branching string net with graphical rules given by Fig. 26.34 (Compare to the problem on Fibonacci pentagon relation) where $\phi^{-1} = (\sqrt{5}-1)/2$. In the ground state no endpoints of strings are allowed, but branching is allowed.

To complete the graphical rules we must also use the rules shown in Fig. 26.35 for some values of the variables, d, X and T.

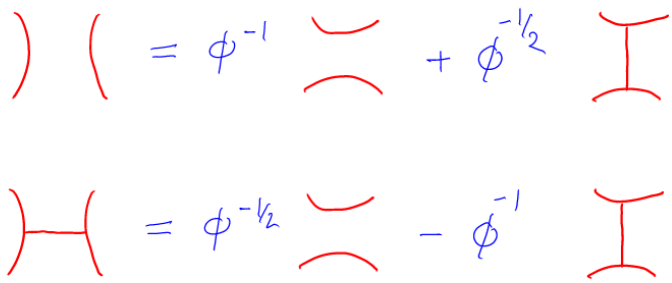


Fig. 26.34 String net rules for the doubled Fibonacci model

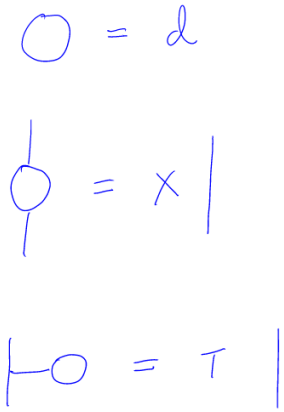


Fig. 26.35 Additionnal string net rules for the doubled Fibonacci model

(a) Show that the consistent solutions is $d = \phi$ with $X = \phi^{1/2}$ and $T = 0$. We did much of this in lecture. What was left out is proving that any $T \neq 0$ solution is not self-consistent. Hint: Try evaluating a circle with three legs coming out of it. That should enable you to derive a useful identity. Then see if you can use this identity to derive a contradiction when $T \neq 0$.

(b) Consider quasiparticles which are the ends of strings. The general form of a quasiparticle is as shown in Fig 26.36 with coefficients a, b, c that need to be determined. Find the eigenvalues/eigenvectors of the rotation operator to determine the quasiparticle types and their spins. (We did most of this in lecture except the explicit evaluation of the eigenvalue problem!) Compare your result to the result of the problem "Fibonacci Hexagon Equation".

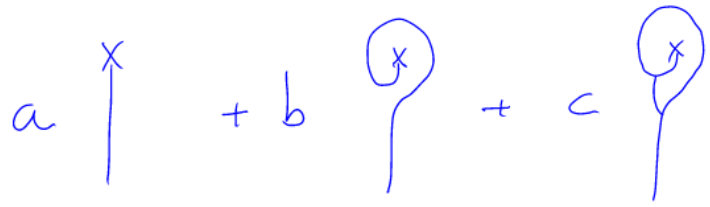


Fig. 26.36 Combination of defect types for the doubled Fibonacci model

Introduction to Quantum Hall — The Integer Effect

27

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.

[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.

27.1 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. 27.1). 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.

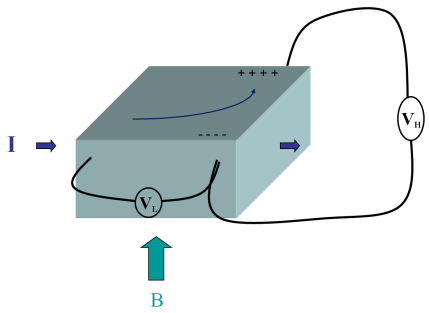


Fig. 27.1 Hall voltage V_H perpendicular to both magnetic field and current, and proportional to both. Also one measures a longitudinal voltage in the same direction as the current, roughly independent of magnetic field.

27.2 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

[2]Metal Oxide Semiconductor Field Effect Transistors

[3]More recently people have been able to produce materials like graphene which are literally one atom thick!

MOSFETs[2] and later quantum wells were used to provide a confining potential for electrons in one direction[3], 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. 27.2.

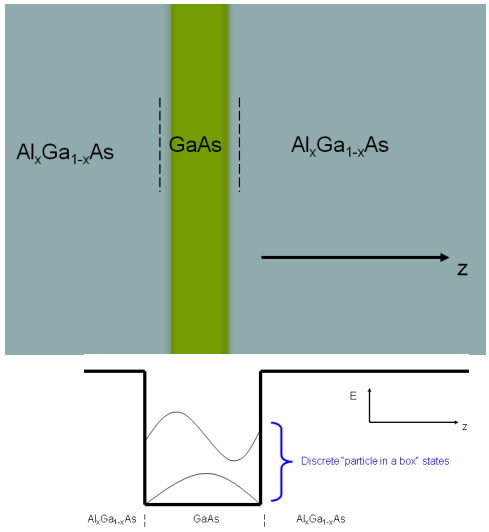


Fig. 27.2 **Top** A quantum well structure is a quasi-two-dimensional layer of one semiconductor sandwiched between two other semiconductors. **Bottom** The potential felt by an electron is like a particle in a box. If the energy is low enough, the electron is stuck in the lowest particle-in-box wavefunction $\varphi_0(z)$ 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. 27.2. 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).

27.3 Phenomenology of Integer Quantum Hall Effect

In 1980 Klaus von Klitzing, having just left a postdoctoral position at Oxford, went to a new job at Grenoble carrying some new high mobility[4] two dimensional electron samples grown by (now Sir) Michael Pepper at Cambridge. He put them in high magnetic field and cooled them down to a few degrees Kelvin temperature where he discovered something very different from what Hall had seen a hundred years earlier. An example of this type of experiment is shown in Fig. 27.3.

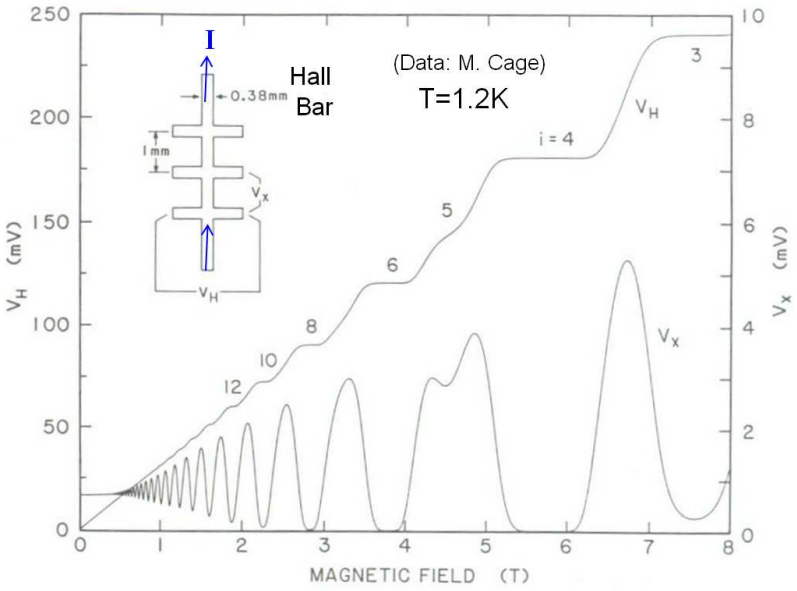


Fig. 27.3 An example of an Integer Quantum Hall experiment. The plateaus in V_H are such that $V_H = (1/i)(h/e^2)I$ with i the integer displayed over the plateau — where h is Planck's constant and e is the electron charge. At the same magnetic field where a plateau occurs in V_H the longitudinal voltage 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

$$V_H = \frac{1}{i}\frac{h}{e^2}I$$

[4]Meaning very clean

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

$$R_H = \frac{1}{i}\frac{h}{e^2} = 1/G_H \tag{27.1}$$

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. 27.1 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.

[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 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.

27.4 Transport in Zero Disorder

In strictly zero disorder it is easy to show that the longitudinal resistance is zero and the Hall resistance is precisely linear in the magnetic field. This is a simple result of Galilean/Lorentz invariance. Suppose we have a two dimensional disorder-free system of electrons in the x, y plane and a magnetic field $\mathbf{B} = B\hat{\mathbf{z}}$ in the $\hat{z}$-direction perpendicular to the plane.

The Lorentz force on an electron will be

$$\mathbf{F} = -e\left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right)$$

If we then boost into a moving frame where

$$\mathbf{v} = \frac{\mathbf{E} \times \hat{z}}{|B|}$$

in this new frame we obtain $\mathbf{F} = \mathbf{0}$, so the ground state must be stationary in this frame.

Then we boost back into the lab frame, and we obtain a current

$$\mathbf{j} = -en\mathbf{v} = \frac{-en\mathbf{E} \times \hat{z}}{|B|}$$

thus giving us

$$\begin{aligned} R_L &= 0 \\ R_H &= \frac{B}{ne} \end{aligned}$$

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.

27.5 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

$$\left(\frac{p_x^2}{2m} + \frac{1}{2}m\omega_c^2(k_y\ell^2 + x)^2\right)\phi_{k_y}(x) = E\phi_{k_y}(x) \tag{27.2}$$

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

$$E_p = \hbar\omega_c\left(p + \frac{1}{2}\right) \tag{27.3}$$

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. 27.4.

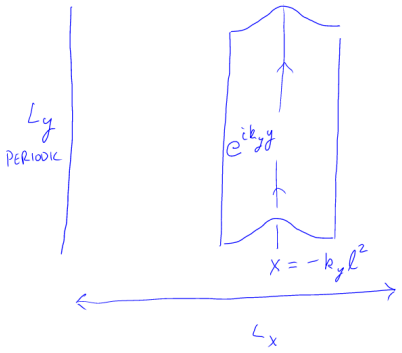


Fig. 27.4 The shape of the wavefunction of an electron in a magnetic field using Landau gauge. The form of the wavefunction will be harmonic oscillator on the x direction and plane-wave in the y-direction

Fixing the energy by fixing p in Eq. 27.3, the value of k_y is quantized in units of $2\pi/L_y$. Further, the position x ranges over L_x, meaning that k_y ranges over L_x/ℓ^2. Thus the total number of possible values of k_y is

$$\text{Number of states in a Landau level} = \frac{L_x L_y}{2\pi\ell^2} = \frac{\text{Area} \quad B}{\phi_0}$$

where

$$\phi_0 = h/e$$

is the magnetic flux quantum. Thus, the number of states in a Landau level is equal to the number of magnetic flux quanta of magnetic field incident on the plane.

We can plot the density of states for electrons in a magnetic field, as shown in Fig. 27.5

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

$$N_e = \nu N_\phi$$

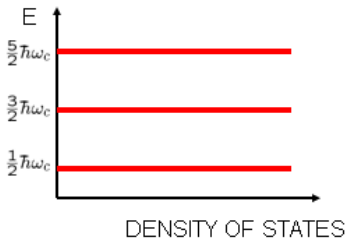


Fig. 27.5 The density of states for spin-polarized (or spinless) electrons in a magnetic field. At energies equal to half-odd integer multiples of the cyclotron frequency, there is a spike of degenerate states, with degeneracy $\frac{\text{Area} \quad B}{\phi_0}$.

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. 27.6.

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. 27.6.

When the chemical potential is anywhere in the localized states, then at low enough temperature, the electrons cannot move at all. Although

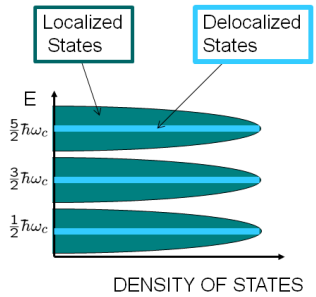


Fig. 27.6 The density of states for spin-polarized (or spinless) electrons in a magnetic field with disoder. The Landau bands are spread out, with localized eigenstates in the tails and extended eigenstates near the middle.

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.

27.6 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 chapter ***.

27.6.1 Byers and Yang Theorem

Consider any system (made of electrons and protons and neutrons) with a hole cut in it, as in Fig. 27.7. 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).

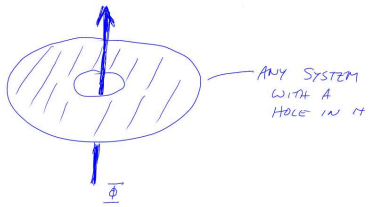


Fig. 27.7 The Byers-Yang theorem states that threading any integer number of flux quanta through a hole in a system leaves the eigenspectrum unchanged.

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, 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

[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.

enclosed

$$\Phi' = \oint \mathbf{A}' \cdot \mathbf{dl} = \oint \mathbf{A} \cdot \mathbf{dl} + 2\pi m\hbar/e = \Phi + m\phi_0$$

has changed by an integer number of flux quanta.

27.6.2 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. 27.8. 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 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

$$J = G_H \mathcal{E}$$

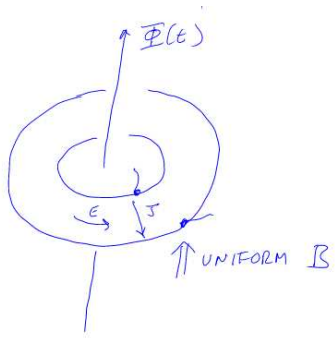


Fig. 27.8 Insertion of Flux $\Phi(t)$ through the center of an annulus of two-dimensional electrons in a uniform magnetic field. Adiabatically increasing the flux creates an electric field in the annular direction which then, by the Hall conductivity, creates current in the radial direction.

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

[11] For studying current flow in magnetic fields, the annulus is knowni as "Corbino" geometry, after O. M. Corbino, who studied this in 1911.

[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.

of "spectroscopy" to measure the charge on the electron! (hence the precision of the effect).

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. We will return to this issue in section 27.8 below.

27.7 Edge States

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. 27.9. 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.

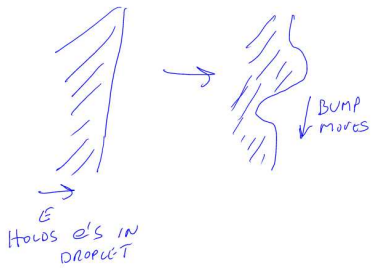


Fig. 27.9 A deformation of the edge is a low energy edge excitation which moves along the edge due to $E \times B$ drift.

27.7.1 Landau Gauge Edge Picture for Integer Quantum Hall

Recall in Landau gauge (See section 27.5) 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. 27.10.

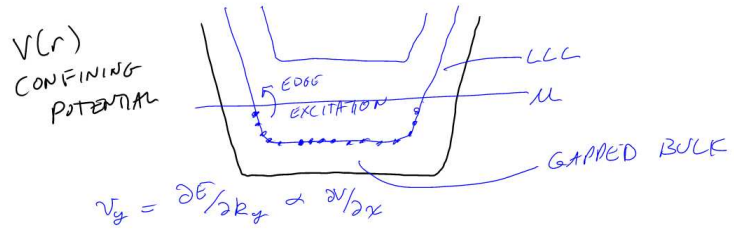


Fig. 27.10 Low energy edge excitations

The addition of the confining potential $V(x)$ simply adds this potential to the 1-d schroedinger equation 27.2. 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. 27.10.

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 $\hbar\Delta k_y$. [13] 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.

27.8 The Halperin Refinement of Laughlin's Argument

A more careful version of Laughlin's 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. 27.11.

[13] 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. ($\Delta\mu = -e\Delta V$).

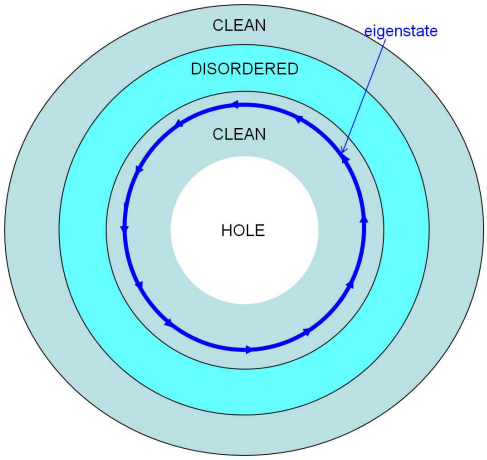


Fig. 27.11 The Halperin geometry. The same as the Laughlin annulus geometry, except here we add disorder only in part of the annulus. We have also shown (dark blue) a single particle eigenstate in the clean region, which forms a circle (with a small gaussian cross-section).

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[14] 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.

[14] Just find the maximum of $|\psi_m|^2$.

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. The only low energy excitations are the edge states!

Now, let us track what happens to the eigenstates as we change the flux through the hole. If the flux through the hole is an integer (in units of the flux quantum ϕ_0), then the angular momentum is also an integer. However, if the flux through the hole is an integer plus some fraction α, then the angular momentum quantum number must also be an integer plus α. Thus, as we adiabatically increase the flux by one flux quantum, we adiabatically turn each m eigenstate to $m + 1$. Thus

we are continuously pushing out electrons to the next further out radial wavefunction.

Now when we are in the disordered region of the annulus, we do not know any details of the shape of the eigenstates. All we know is that after insertion of a full flux quantum we must get back to the same many body eigenstate that we started with. However, we also know that an additional electron is being pushed into the disordered region from the clean region on the inside, whereas an electron is also being extracted into the clean region on the outside. Thus the disordered region must also convey exactly one electron (per Landau level) when a flux quantum is inserted adiabatically. An electron state is moved from one edge state on the inside to an edge state on the outside.

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).

Exercises

Exercise 27.1 *Quantum Hall Conductivity vs Conductance*

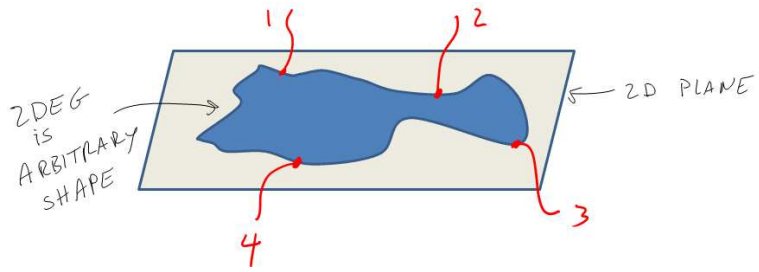


Fig. 27.12 A 2D electron gas of arbitrary shape with contacts 1,2,3,4 attached on its perimeter in clockwise order

Consider a two dimensional electron gas (2DEG) of arbitrary shape in the plane with four contacts (1,2,3,4) attached at its perimeter in a clockwise order as shown in Fig. 27.12. The conductivity tensor σ_{ij} relates the electric field to the current via

$$j_i = \sigma_{ij} E_j \tag{27.4}$$

where indices i and j take values $\hat{x}$ and $\hat{y}$ (and sum over j is implied). Assume that this is a quantized hall system with quantized hall conductance s. In other words, assume that

$$\sigma = \begin{pmatrix} 0 & s \\ -s & 0 \end{pmatrix} \tag{27.5}$$

Show that the following two statements are true independent of the shape of the sample.

(a) Suppose current I is run from contact 1 to contact 2, show that the voltage measured between contact 3 and 4 is zero.

(b) Suppose current I is run from contact 1 to contact 3, show that the

voltage measured between contact 2 and 4 is $V = I/s$.

Note: The physical measurements proposed here measure the *conductance* of the sample, the microscopic quantity σ is the *conductivity*.

Exercise 27.2 *About the Lowest Landau Level*

If you have never before actually solved the problem of an electron in two dimensions in a magnetic field, it is worth doing. Even if you have done it before, it is worth doing again.

Consider a two dimensional plane with a perpendicular magnetic field $\vec{B}$. Work in symmetric gauge $\vec{A} = \frac{1}{2}\vec{r} \times \vec{B}$.

(a) (This is the hard part, see below for hints if you need them.) Show that the single electron Hamiltonian can be rewritten as

$$H = \hbar\omega_c(a^\dagger a \ + \frac{1}{2}) \tag{27.6}$$

where $\omega_c = eB/m$ and

$$a = \sqrt{2}\ell\left(\bar{\partial} + \frac{1}{4\ell^2}z\right) \tag{27.7}$$

with $z = x + iy$ and $\bar{\partial} = \partial/\partial\bar{z}$ with the overbar meaning complex conjugation. Here ℓ is the magnetic length $\ell = \sqrt{\hbar/eB}$.

(b) Confirm that

$$[a, a^\dagger] = 1 \tag{27.8}$$

and therefore that the energy spectrum is that of the harmonic oscillator

$$E_n = \hbar\omega_c(n + \frac{1}{2}) \tag{27.9}$$

(c) Once you obtain Eq. 27.6, show that any wavefunction

$$\psi = f(z)e^{-|z|^2/4\ell^2} \tag{27.10}$$

with f any analytic function is an eigenstate with energy $E_0 = \frac{1}{2}\hbar\omega_c$. Show that an orthogonal basis of wavefunctions in the lowest Landau level (i.e., with eigenenergy E_0) is given by

$$\psi_m = N_m z^m e^{-|z|^2/4\ell^2} \tag{27.11}$$

where N_m is a normalization constant. Show that the maximum amplitude of the wavefunction ψ_m is a ring of radius $|z| = \ell\sqrt{2m}$ and calculate roughly how the amplitude of the wavefunction decays as the radius is changed away from this value.

(d) Defining further

$$b = \sqrt{2}\ell\left(\partial + \frac{1}{4\ell^2}\bar{z}\right) \tag{27.12}$$

with $\partial = \partial/\partial z$, Show that the operator b also has canonical commutations

$$[b, b^\dagger] = 1 \tag{27.13}$$

but both b and $b^\dagger$ commute with a and $a^\dagger$. Conclude that applying b or $b^\dagger$ to a wavefunction does not change the energy of the wavefunction.

(e) show that the $\hat{z}$ component of angular momentum (angular momentum perpendicular to the plane) is given by

$$L = \hat{z} \cdot (\vec{r} \times \vec{p}) = \hbar(b^\dagger b \ - a^\dagger a \ ) \tag{27.14}$$

Conclude that applying b or $b^\dagger$ to a wavefunction changes its angular momentum, but not its energy.

(f) [Harder] Let us write an arbitrary wavefunction (not necessarily lowest Landau level) as a polynomial in z and $\bar{z}$, times the usual gaussian factor. Show that projection of this wavefunction to the lowest Landau level can be performed by moving all of the $\bar{z}$ factors all the way to the left and replacing each $\bar{z}$ with $2\ell^2\partial_z$.

Hints to part a: First, define the antisymmetric tensor ϵ_{ij}, so that the vector potential may be written as $A_i = \frac{1}{2}B\epsilon_{ij}r_j$. We have variables p_i and r_i that have canonical commutations (four scalar variables total). It is useful to work with a new basis of variables. Consider the coordinates

$$\pi_i^{(\alpha)} = p_i + \alpha\frac{\hbar}{2\ell^2}\epsilon_{ij}r_i \tag{27.15}$$

$$= \frac{\hbar}{\ell^2}\epsilon_{ij}\xi_j \tag{27.16}$$

defined for $\alpha = \pm 1$. Here $\alpha = +1$ gives the canonical momentum. Show that

$$\left[\pi_i^{(\alpha)}, \pi_j^{(\beta)}\right] = i\alpha\epsilon_{ij}\delta_{\alpha\beta}\frac{\hbar^2}{\ell^2} \tag{27.17}$$

The Hamiltonian

$$H = \frac{1}{2m}(p_i + eA_i)(p_i + eA_i) \tag{27.18}$$

can then be rewritten as

$$H = \frac{1}{2m}\pi_i^{(+1)}\pi_i^{(+1)} \tag{27.19}$$

with a sum on $i = \hat{x}, \hat{y}$ implied. Finally use

$$a = (-\pi_y^{(+1)} + i\pi_x^{(+1)})\frac{\ell}{\sqrt{2}\hbar} \tag{27.20}$$

$$b = (\pi_y^{(-1)} + i\pi_x^{(-1)})\frac{\ell}{\sqrt{2}\hbar} \tag{27.21}$$

to confirm that a and b are given by Eqs. 27.7 and 27.12 respectively. Finally confirm Eq. 27.6 by rewriting Eq. 27.19 using Eqs. 27.20 and 27.21.

A typical Place to get confused is the definition of ∂. Note that

$$\partial z = \bar{\partial}\bar{z} = 1 \tag{27.22}$$

$$\bar{\partial} z = \partial\bar{z} = 0 \tag{27.23}$$

Hints to part f: Rewrite the operators $a, a^\dagger, b, b^\dagger$ such that they operate on polynomials, but not on the Gaussian factor. Construct $\bar{z}$ in terms of these operators. Then project.

Aside: A Rapid Introduction to Topological Insulators

28

The integer quantum Hall effect is one of the simplest examples of what is now called a "topological insulator". To explain what this is, and why it is interesting, let us review some basic facts about band structure and non-interacting electrons.[1]

[1] In this chapter we are thinking about non-interacting electrons in periodic potentials!

28.1 Topological Phases of Matter

We will consider systems of electrons in some periodic environment — which is what an electron would experience in a real material crystal[2]. We can thus describe our system as some single electron kinetic energy and some periodic potential — or equivalently as some tight-binding model. Bloch's theorem tells us that the eigenstates of such a periodic Hamiltonian can be written in the form

$$|\Psi_{\mathbf{k}}^{\alpha}\rangle = e^{i\mathbf{k}\cdot\mathbf{r}}|u_{\mathbf{k}}^{\alpha}\rangle$$

where α is the band index, and $u_{\mathbf{k}}^{\alpha}(\mathbf{x})$ is a function periodic in the unit cell.

[2] Some of the ideas discussed here do not depend too much on the system being precisely periodic.

The eigen-spectrum breaks up into bands of electron states. If a (valence) band is completely filled and there is a gap to next (conduction) band which is empty, we generally call the system a band insulator. The conventional wisdom in most solid state physics books is that such band insulators carry no current. This wisdom, however, is not correct. A prime example of this is the integer quantum hall effect! As we have just seen for the integer quantum Hall effect we have a filled band and a gap in the single electron spectrum. And while such a system carries no longitudinal current (and correspondingly has $\sigma_{xx} = 0$) it does carry Hall current with $\sigma_{xy} = ne^2/h$.

One might object that the integer quantum Hall effect is not really a valid example, because it does not have a periodic potential. However, it is certainly possible to add a very weak periodic potential to the quantum Hall system and maintain the gap.

It turns out that there is a topological distinction in the wavefunctions for the quantum Hall effect versus what we think of as a traditional band insulator. One way to describe this is to think of the band structure as being a mapping from the Brillouin zone (inequivalent values of $\mathbf{k}$) to

the space of possible wavefunctions

$$\mathbf{k} \to u^{\alpha}_{\mathbf{k}}(\mathbf{x}). \tag{28.1}$$

Once we have such a mapping we can ask about whether there are topologically different mappings, or whether one mapping can be continuously deformed to another.

An analogy is to consider a mapping from a circle S^1 to a circle S^1,

$$e^{i\theta} \to e^{if(\theta)}$$

Here, one can topologically classify the mapping by its winding number. One such mapping cannot be continuously deformed into another if the two mappings have different winding numbers.

Similarly we can define a "winding number" (known as a "Chern" number) of the band structure map Eq. 28.1 for two dimensional systems. This integer topological quantity turns out to be precisely the quantized Hall conductance in units of e^2/h. We give an explicit expression for this quantity in section *** below. Similar topological definitions of "winding numbers" of the map Eq. 28.1 can be given in any dimension.

If we imagine continously changing the physical Hamiltonian, this Chern number, which must be an integer, cannot change continously. It can only change by making it impossible to define a Chern number. This happens when if the system becomes a metal — i.e, if the gap between the filled and empty state closes. Thus we cannot deform between different topological classes without closing the gap.

Indeed, this general picture gives us a simple rule for topological classification:

> **Definition of Topological Phase:** *Two gapped states of matter are in the same topological phase of matter if and only if you can continuously deform the Hamiltonian to get from one state to the other without closing the excitation gap.*

Although in this chapter we are concerned with non-interacting electrons only, this sort of definition can obviously be used much more generally to distinguish different phases of matter. Further this definition fits with our intuition about topology

> *Two objects are topologically equivalent if and only if you can continuously deform one to the other.*

In the context of noninteracting electron band structure, one can define topologically "trivial" phases of matter to be those that can be continuously deformed without closing the gap into individual atomic sites with electrons that do not hop between sites. (A "trivial" band structure). Phases of matter that cannot be continuously deformed to this trivial band structure without closing a gap are known as topologically nontrivial.

28.1.1 Gapless Edges

The existence of gapless edge states on the edge of integer quantum Hall samples is one of the fundamental properties of topologically nontrivial phases of matter (at least when one is considering topological properties of noninteracting electron band structure). We can give a rough argument about why edge states always come with topologically nontrivial phases.

Suppose we have a Hamiltonian that is almost periodic, but the potential is a very function of position, say in the x-direction. In other words if we move very far in the x-direction the Hamiltonian changes smoothly from $H(x_1)$ to $H(x_2)$, but locally both of these look like simple periodic Hamiltonians. If $H(x_1)$ and $H(x_2)$ are not in the same topological phase of matter, than for some x between x_1 and x_2, we have $H(x)$ describing some gapless system — i.e., an edge state between the two phases.

For example, in the case of the integer quantum Hall effect, we can think of $H(x_1)$ as being the Hamiltonian of the system in the bulk which has nonzero Chern number, and $H(x_2)$ as being the Hamiltonian outside of the system, or the vacuum, which is topologically trivial and has zero Chern number. Somewhere between the two, the gap must close to give a metal where the Chern number changes. This is the edge state.

28.2 Curvature and Chern Number

The Gauss-Bonnet theorem give an beautiful connection exists between topology and geometry. The statement of the theorem is that for any closed two dimensional orientable surface the integral of the Gaussian curvature K over the surface gives $2\pi(2-2g)$ where g is the number of handles of the surface. Or mathematically[3]

$$2\pi(2-2g) = \int_M K dS$$

[3]The definition of Gaussian curvature K at a point is $1/K = \pm r_{max} r_{min}$ where r_{max} and r_{min} are the maximum and minimum radii of curvatures of the surface at that point. The sign of K is taken to be negative if the surface is saddle-like at that point rather than dome-like.

One can check, for example, with a sphere of radius R we have $K = 1/R^2$ and $g = 0$, so that both sides give 4π independent of R. The interesting point here is that if you dent the sphere, you increase the curvature at some points, but you decrease it at other points such that the integral of the curvature over the surface remains the same. The only way to change this quantity is to rip the surface and add a handle!

It turns out that we can define a similar curvature that describes the topological index (the Chern-number) of the band structure. Let us define what is known as the Berry curvature of the α^{th} band

$$\mathcal{F}^\alpha(\mathbf{k}) = \epsilon^{ij} \langle \partial_{k_i} u^\alpha_{\mathbf{k}} | \partial_{k_j} u^\alpha_{\mathbf{k}} \rangle$$

The topological Chern-number of the α^{th} filled band is then given by

the integral of the Berry curvature over the Brillouin zone,

$$C^\alpha = \frac{1}{2\pi}\int_{BZ} d\mathbf{k}\ \mathcal{F}^\alpha(\mathbf{k})$$

which is analogously quantized to be an integer.

In appendix *** we use the Kubo formula to calculate the Hall conductivity and we find that it is related to the Chern number by[4]

$$\sigma_{xy} = \frac{e^2}{h}\sum_{\text{filled bands }\alpha} C^\alpha$$

Considering Laughlin's proof that the Hall conductance is quantized, this might be considered a sufficient proof that the Chern number must be quantized as well. To see how this occurs mathematically, see appendix ***.

[4]The realization that the Hall conductance is the topological Chern number in 1982 was made in a famous paper known as TKNN. This is one of key contributions that earned a Nobel Prize for David Thouless in 2016.

28.3 Symmetry Protection

Symmetry is one of the most fundamental ideas in modern physics. We often think about how physics changes when a symmetry is forced on a system. Considering the above definition of topological phases of matter in section 28.1, one may generalize this idea to systems with symmetry.

Definition of Symmetry Protected Topological Phase: *Two gapped states of matter are in the same symmetry protected topological phase of matter if and only if you can continuously deform the Hamiltonian to get from one state to the other without closing the excitation gap or breaking the given symmetry.*

The most interesting example of this is time reversal symmetry. Systems without magnetism and without magnetic impurities are time-reversal symmetric. In three dimensions, it turns out that there are no band structures that satisfy the above definition of a nontrivial topological phase of matter. In other words, all gapped periodic single-electron Hamiltonians can be deformd to a trivial Hamiltonian without closing the gap. However, if we enforce time reversal invariance, it turns out that there *are* band structures that cannot be deformed into the trivial band structure without closing the gap or breaking symmetry. These are known as "topological insulators" and are formally symmetry protected topological phases, where the symmetry is time reversal.

28.4 Appendix: Chern Number is Hall Conductivity

Here we calculate the Hall conductivity by simple time dependent perturbation theory and demonstrate that it is the same as the Chern number.

The general rule of time dependent perturbation theory is that if a system is exposed to a perturbation $\delta H(t)$ the expectation of an operator O at some later time is given by

$$\langle O(t)\rangle = \frac{i}{\hbar}\int_{-\infty}^{t} dt' \langle [O(t), H(t')]\rangle$$

If we consider an electric field at frequency ω we write this in terms of the vector potential. Applying a perturbing vector potential we have

$$\delta H = \int d\mathbf{x}\, \mathbf{A}(\mathbf{x},\mathbf{t})\cdot \mathbf{j}(\mathbf{x},\mathbf{t})$$

From perturbation theory we then have

$$\langle j_a(x,t)\rangle = \frac{i}{\hbar}\int_{-\infty}^{t} dt' \int d\mathbf{x}' \langle [j_a(x,t), j_b(x',t')] A_b(x',t')$$

Introduction to Fractional Quantum Hall Effect

29

charge-flux

Having determined that the quantum Hall effect is some sort of spectroscopy on the charge of the electron, it was particularly surprising in 1982 when Dan Tsui and Horst Stormer[1] discovered quantum Hall plateaus at fractional values of the filling fraction

$$\nu = p/q$$

with Hall resistance

$$R_H = \frac{h}{e^2}\frac{q}{p}$$

with p and q small integers. This effect is appropriately called the Fractional quantum Hall effect.

The first plateau observed was the $\nu = 1/3$ plateau[2] , but soon thereafter many more plateaus were discovered[3]. The Nobel Prize for this discovery was awarded in 1998.

Given our prior gauge invariance argument that quantum Hall effect is measuring the charge of the electron — and that this is enforced by the principle of gauge invariance, it is hard to understand how the fractional effect can get around our prior calculation.

Two things must be true in order to have quantized Hall effect

> (a) Charge must fractionalize into quasiparticles with charge $e^* = e/q$, for example in the case of $\nu = 1/q$.
>
> (b) The ground state on an annulus must be degenerate, with q different ground states (in the case of $\nu = 1/q$) which cycle into each other by flux insertion through the annulus.

We should not lose sight of the fact that these things are surprising — even though the idea of degenerate ground states, and possibly even fractionalized charges, is something we have perhaps gotten used to in our studies of topological systems.

Given the Laughlin argument that inserting a flux though the annulus pumps an integer number of electrons from one side to the other, it is perhaps not surprising that fractional quantization of the Hall conductance must imply that a *fractional* charge has been pumped from one side of the annulus to the other (hence point (a) above). The way we get around the gauge invariance argument that implies the charge must

[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!

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

$$\begin{pmatrix} 3N_e \\ N_e \end{pmatrix} \sim (27/4)^{N_e}$$

multiparticle states to choose from — and all of these states have the same energy! In the thermodynamic limit this is an insanely enormous degeneracy[4]. This enormous degeneracy is broken by the interaction between the electrons, which will pick out a very small ground state manifold (in this case being just 3 degenerate ground states), and will leave the rest of this enormous Hilbert space with higher energy.

[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.

29.0.1 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 27.5, 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. 27.5, 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

$$H = \sum_{i<j} V(\mathbf{r}_i - \mathbf{r}_j) \tag{29.1}$$

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!

29.1 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 27.8) take the form[7]

$$\varphi_m(z) = C_m z^m e^{-|z|^2/(4\ell^2)} \tag{29.2}$$

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 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. 29.2. 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.

29.1.1 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[8]

$$\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N) = (\text{Polynomial in } z_1, \ldots z_N) \prod_{i=1}^{N} e^{-|z_i|^2/(4\ell^2)}$$

[7] We will ignore dom as before.

[8] The polynomial can also be chosen so as to have all real coefficients. This is becuase the Hamiltonian, once projected to a single Landau level, i.e., Eq. 29.1, is time reversal symmetric.

(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![9]) to consider fractional quantum Hall effect of bosons as well — in which case the wavefunction should be fully symmetric.

[9] While no one has yet produced fractional quantum Hall effect of bosons in the laboratory, 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.

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!

29.2 Laughlin's Ansatz

Laughlin simply guessed that a good wavefunction would be of the form[10]

$$\Psi^{(m)}_{Laughlin} = \prod_{i<j} (z_i - z_j)^m \prod_{i=1}^{N} e^{-|z_i|^2/(4\ell^2)}$$

[10] Note that this wavefunction is not normalized in any sense. The issue of normalization becomes important later in ***.

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 wavefunction 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)[11].

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 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.

29.2.1 Exact statements about Laughlin Wavefunction

It turns out that the Laughlin wavefunciton is actually the exact ground state of a special inter-particle interaction[12].

[12] This was discovered by Haldane in 1983, then again by Trugman and Kivelson and also Pokrovski and Talapov in 1985.

Bosons at $\nu = 1/2$

Consider a system of bosons with the interparticle interaction given by[13]

$$V = V_0 \sum_{i<j} \delta(\mathbf{r}_i - \mathbf{r}_j)$$

with $V_0 > 0$. This is a non-negative definite interaction.

[13] Actually this is a very realistic interaction for cold atom bosonic quantum Hall effect, should it be produced in the future.

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[14]. 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

[14] Although with some thought this fact seems obvious, proving it rigorously is tricky.

[11] 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.

less dense. A schematic of the ground state energy as a function of filling fraction for this case is shown in Fig. 29.1.

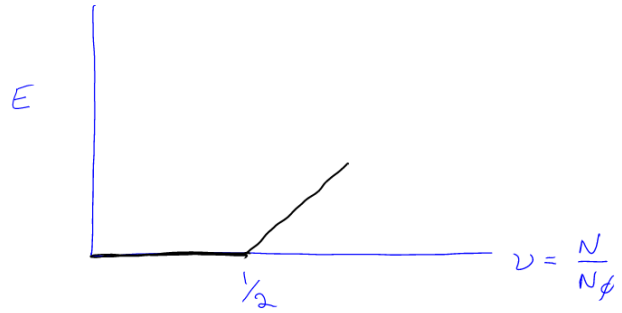


Fig. 29.1 Schematic of the ground state energy as a function of filling fraction for bosons with delta function interaction.

The key point is that the ground state energy has a cusp, which means there is a jump in the chemical potential

$$\mu = \frac{\partial E}{\partial N}$$

This is precisely the same "incompressibility" as we have in the case of noninteracting electrons — where the chemical potential jumps between Landau levels! As in that case we presume that the presence of a cusp in the free energy, in the absence of disorder, will be enough to give us a plateau when disorder is added back in.

Now while we can easily show that there is a change of behavior at $\nu = 1/2$ in this plot, it is somewhat more difficult to be convincing that the slope coming from the right is finite — i.e., that the gap is actually finite. In order to do that, we would need to think about the elementary excitations, or resort to numerics.

Fermions at $\nu = 1/3$

The arguments given for bosons at $\nu = 1/2$ can be easily generalized to the case of fermions (i..e, electrons) at $\nu = 1/3$ (and more generally to any $\nu = 1/m$.) Obviously a δ-function interaction will no longer do the job, since for fermions Pauli exclusion prevents any two fermions from coming to the same point already. However, consider an interaction of the form

$$V = V_0 \sum_{i<j} \nabla^2 \delta(\mathbf{r}_i - \mathbf{r}_j)$$

Given a wavefunction $\Psi(r_1, \ldots, r_N)$ the interaction energy will be

$$E = \sum_{i<j} \int \mathbf{dr}_1 \ldots \mathbf{dr}_N \; |\Psi|^2 \;\; \nabla^2 \delta(\mathbf{r}_i - \mathbf{r}_j)$$

Writing

$$\Psi(\mathbf{dr}_1 \ldots \mathbf{dr}_N) = \phi(z_1 \ldots z_N) \prod_{i=1}^{N} e^{-|z_i|^2/(4\ell^2)} \tag{29.3}$$

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[15] 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 powers $\partial_z \phi$ will remain zero (since $\partial_{z_i}(z_i - z_j)^3$ goes to zero when $z_i = z_j$)[16].

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.

29.2.2 Real Interactions

Obviously electrons do not interact via a $\nabla^2\delta$ interaction. They interact via a Coulomb interaction[17] 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[18] 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! The energy of the Laughlin wavefunction differs from the energy of the exact Coulomb ground state by less than a part in two thousand[19].

29.3 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

[15] 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.

[16] Note that by antisymmetry the wavefunction must vanish as an odd number of powers as two particle positions approach each other.

[17] In higher Landau levels, although the interaction is Coulomb, when the single Landau level problem is mapped to a single partly filled *lowest* Landau level (See the comments after Eq. 29.2), the interaction gets modified – this mainly effects the short range behavior.

[18] The full Hilbert space is 45207 dimensional!

[19] I need to recheck this number***.

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.*

29.3.1 Quasiholes

For the quasiholes, it is fairly easy to guess their wavefunction (and indeed this was done by Laughlin). We start by considering adding a quasihole at postion $\mathbf{0}$. This leaves the system rotationally invariant. We guess the solution

$$\Psi_{qh}(\mathbf{0}) = \left[\prod_{i=1}^{N} z_i\right] \Psi_{Laughlin}$$

where $\mathbf{0}$ indicates we have put the quasihole at position $\mathbf{0}$. Here the degree of the polynomial is increased by one for every variable, so each filled orbital gets pushed out to the next orbital. This leaves precisely one empty orbtial open at positon $\mathbf{0}$. Since our wavefunction has filling fraction ν, this means that on average a fraction ν of the orbitals are filled. Thus leaving the orbital at the center completely empty corresponds to a positive charge of $+\nu$, and our quasihole has a positive charge

$$e^* = \nu e.$$

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 once we have inserted the flux, the 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 quasiholes 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 is the case for the Laughlin ground state, the quasihole is not exact for the Coulomb interaction, but is extremely accurate numerically.

(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).

29.3.2 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 a 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. 29.3 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

$$\Psi_{qes}(w) = \left(\left[\prod_{i=1}^{N} \left(2\ell^2 \frac{\partial}{\partial z_i} - w^* \right) \right] \phi \right) \prod_{i=1}^{N} e^{-|z_i|^2/(4\ell^2)}$$

29.3.3 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 quasiparticles 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. 29.2

Fig. 29.2 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.

29.4 Digression on Berry's Phase

The Berry phase[20] 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

[20] Berry's work on Berry Phase in 1984 had a number of precursors, most notably the work of Pancharatnam in 1956.

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 of 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 into 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

$$\begin{aligned} i\hbar\frac{\partial}{\partial t}|\Psi(t)\rangle &= H(\mathbf{R}(t))|\Psi(t)\rangle \\ \left[-\hbar\dot{\gamma} + i\hbar\frac{\partial}{\partial t}\right]|\psi(\mathbf{R}(t))\rangle &= E(\mathbf{R}(t))|\psi(\mathbf{R}(t))\rangle \end{aligned}$$

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.

29.5 Arovas-Schrieffer-Wilczek Calculation of Fractional Statistics

This section follows the approach of Arovas, Schrieffer and Wilczek[21].

[21] 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.

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. 29.2. 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[22] the position of the particle by the angle θ and $w = |w|e^{i\theta}$.

[22] 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 ***

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

$$\partial_\theta |\Psi(\theta)\rangle = \frac{\partial w}{\partial \theta} \left(\sum_i \frac{-1}{z_i - w} \right) |\Psi(\theta)\rangle$$

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[23]

$$\begin{aligned} \Delta\gamma &= -i \oint d\theta \, \langle \Psi(\theta) | \partial_\theta | \Psi(\theta) \rangle \\ &= -i \oint dw \sum_i \left\langle \Psi(w) \left| \frac{-1}{z_i - w} \right| \Psi(w) \right\rangle \end{aligned}$$

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. 29.2. Using a normalization factor that is again a function of $|w|$ only, the same calculation holds, but now the

[23] The way this is written it is obviously a bit nonsense. Please fix it. I wrote this footnote, but now I don't see what is wrong with what I have here! ***

number of electrons enclosed has changed by one quasiparticle charge e/m. Thus the phase is now

$$\Delta\gamma = \gamma_{AB} + \gamma_{statistical}$$

where the 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 we 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 charge-flux model). We also have that the fusion of m elementary quasi-electrons or quasi-holes forms an an electron or anti-electron.

In the case where m is even, the underlying "electron" is a boson, in which case we can think of this electron as being identical to the vacuum — it has trivial braiding with all particles and it is essentially condensed into the ground state as some sort of background superfluid. Thus we have a simple anyon theory with m particle types.

On the other hand, when m is odd, we have the situation (discussed in our "charge-flux composite" section ***) where the fusion of m elementary anyons forms a fermion — and so there are actually $2m$ particle types — the fermion full-braids trivially with everything, but has fermionic statistics with itself. This situtation is "non-modular" — it does not have as many ground states as it has particle types. There are only m ground states, despite $2m$ particle types.

29.6 Gauge Choice and Monodromy

The Laughlin wavefunction with M quasiholes takes the form

$$\Psi(w_1, \ldots, w_M; z_1, \ldots, z_N) = \mathcal{N}(w_1, \ldots, w_N) \left[\prod_{\alpha=1}^{M}\prod_{i=1}^{N}(z_i - w_\alpha)\right] \Psi^{(m)}_{Laughlin}(z_1, \ldots, z_N) \tag{29.4}$$

where $\mathcal{N}$ is a normalizing factor.

By using a plasma analogy (this is a homework assignment) we find

that the normalization must be of the form

$$|\mathcal{N}(w_1,\ldots,w_M)| = C\prod_{\alpha<\beta}|w_\alpha - w_\beta|^{1/m}\prod_{\alpha=1}^{M} e^{-|w_\alpha|^2/(4\ell^{*2})}$$

where C is some constant and

$$\ell^* = \sqrt{\frac{\hbar}{e^*B}}$$

is the effective magnetic length for a particle of charge $e^* = e/m$. This choice of normalization assures that

$$\langle\Psi(w_1,\ldots,w_M)|\Psi(w_1,\ldots,w_M)\rangle$$

independent of the position of the quasiholes.

Now, we can choose the phase of the factor $\mathcal{N}$ arbitrarily — this is essentially a gauge choice. In the above Arovas, Schrieffer, Wilczek calculation above, we chose the phase to be real. However, this is just a convention. An intersting different convention is to choose

$$\mathcal{N}(w_1,\ldots,w_N) = C\prod_{\alpha<\beta}(w_\alpha - w_\beta)^{1/m}\prod_{\alpha=1}^{M} e^{-|w_\alpha|^2/(4\ell^{*2})} \qquad (29.5)$$

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' 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.

29.6.1 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

$$\begin{aligned}\Psi(w) &= C(w-w')^{1/m} e^{-(|w|^2+|w'|^2)/(4\ell^{*2})} \times \\ &\quad \prod_i (z_i - w)(z_i - w') \prod_{i<j} (z_i - z_j) \prod_i e^{-|z_i|^2/(4\ell^2)}\end{aligned}$$

with C chosen so that Ψ is normalized independent of the quasihole coordinates.[24] 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

$$\frac{\partial}{\partial \tau} = \frac{\partial w}{\partial \tau}\frac{\partial}{\partial w} + \frac{\partial w^*}{\partial \tau}\frac{\partial}{\partial w^*} \tag{29.6}$$

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

$$\Delta\gamma = -i \oint d\tau \langle \Psi(\tau) | \partial_\tau | \Psi(\tau) \rangle = -i \frac{1}{4\ell^{*2}} \oint (dw w^* - dw^* w)$$

where we have used Eq. 29.6. We now use the complex version of Stokes theorem[25] 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!

[24] 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.

[25] The complex version of Stokes is as follows. Using $w = x + iy$

$$\int_{\partial A} (F dw - G dw^*) = 2i \int_A (\partial_{w^*} F + \partial_w G) dx dy$$

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!

29.7 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)

$$\delta n = j^0 = \frac{1}{2\pi m}\delta B$$

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

$$j^\mu = \frac{-1}{2\pi m}\epsilon^{\mu\nu\lambda}\partial_\nu \delta A_\lambda \tag{29.7}$$

We must, of course have charge conservation as well. This is easy to enforce by writing the current in the form

$$j^\mu = \frac{1}{2\pi}\epsilon^{\mu\nu\lambda}\partial_\nu a_\lambda \tag{29.8}$$

which then automatically satisfies

$$\partial_\mu j^\mu = 0$$

In this language, the effective Lagrangian that produces Eq. 29.7 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. 29.8. By differentiating the Lagrangian with respect to a_μ we generate the equations of motion Eq. 29.7.

More here

29.8 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 29.4, 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. 29.4. 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 gener-

ates 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.

Exercises

Exercise 29.1 *Filled Lowest Landau Level*

Show that the filled Lowest Landau level of non-interacting electrons (a single slater determinant) can be written as

$$\Psi_m^0 = \mathcal{N} \prod_{1\le i<j\le N} (z_i - z_j)^1 \prod_{1\le i\le N} e^{-|z_i|^2/4\ell^2} \tag{29.9}$$

with $\mathcal{N}$ some normalization constant. I.e, this is the Laughlin wavefunction with exponent $m = 1$.

Exercise 29.2 *Laughlin Plasma Analogy*

Consider the Laughlin wavefunction for N electrons at positions z_i

$$\Psi_m^0 = \mathcal{N} \prod_{1\le i<j\le N} (z_i - z_j)^m \prod_{1\le i\le N} e^{-|z|^2/4\ell^2} \tag{29.10}$$

with $\mathcal{N}$ a normalization constant. The probability of finding particles at positions $\{z_1, \ldots, z_N\}$ is given by $|\Psi_m(z_1, \ldots z_N)|^2$.

Consider now N classical particles at temperature $\beta = \frac{1}{k_b T}$ in a plane interacting with logarithmic interactions $v(\vec{r}_i - \vec{r}_j)$ such that

$$\beta v(\vec{r}_i - \vec{r}_j) = -2m \log(|\vec{r}_i - \vec{r}_j|) \tag{29.11}$$

in the presence of a background potential u such that

$$\beta u(|\vec{r}|) = |\vec{r}|^2/(2\ell^2) \tag{29.12}$$

Note that this log interaction is "Coulombic" in 2d (i.e., $\nabla^2 v(\vec{r}) \propto \delta(\vec{r})$).

(a) Show that the probability that these classical particles will take positions $\{\vec{r}_1, \ldots, \vec{r}_N\}$ is given by $|\Psi_m^0(z_1, \ldots z_N)|^2$ where $z_j = x_j + iy_j$ is the complex representation of position $\vec{r}_i$. Argue that the mean particle density is constant up to a radius of roughly $\ell\sqrt{Nm}$. (Hint: Note that u is a neutralizing background. What configuration of charge would fully screen this background?)

(b) Now consider the same Laughlin wavefunction, but now with M quasiholes inserted at positions $w_1, \ldots, w_M$.

$$\Psi_m = \mathcal{N}(w_1, \ldots, w_M) \left[\prod_{1\le i\le N} \prod_{1\le \alpha\le M} (z_i - w_\alpha) \right] \Psi_m^0 \tag{29.13}$$

where $\mathcal{N}$ is a normalization constant which may now depend on the positions of the quasiholes. Using the plasma analogy, show that the $w-z$ factor may be obtained by adding additional logarithmically interacting charges at positions w_i,with $1/m$ of the charge of each of the z particles

(c) Note that in this wavefunction the z's are physical parameters (and the wavefunction must be single-valued in z's), but the w's are just parameters of the wavefunction – and so the function $\mathcal{N}$ could be arbitrary — and is only fixed by normalization. Argue using the plasma analogy that in order for the wavefunction to remain normalized (with respect to integration over the z's)

as the w's are varied, we must have

$$|\mathcal{N}(w_1,\ldots,w_M)| = \mathcal{K} \prod_{1\le\alpha<\gamma\le M} |w_\alpha - w_\gamma|^{1/m} \prod_{1\le\alpha\le M} e^{-|w_\alpha|^2/(4m\ell^2)} \quad (29.14)$$

with $\mathcal{K}$ a constant so long as the $w's$ are not too close to each other. (Hint: a plasma will screen a charge).

30 Fractional Quantum Hall Edges

30.1 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[1]

$$\varphi_m \sim z^m e^{-|z|^2/(4\ell^2)}$$

where m is the eigenvalue of the angular momentum[2] operator $\hat{L}$. Since the radius of these states is $r \approx \ell\sqrt{2m}$ it is not surprising that the confinement energy γr^2 of each eigenstate is proportional to m. We thus have

$$H_{confined} = H_0 + \alpha \hat{L}$$

for some constant α.

For integer filling, the edge excitations are very much like the edge excitations we discussed above in Landau gauge. A round quantum Hall droplet fills m states up to a chemical potential along the edge. One can add a small amount of angular momentum to the edge by exciting a filled state from an m just below the chemical potential to an empty state just above the chemical potential.

[1] Note that the parabolic confinement modifies the magnetic length.

[2] We drop the $\hbar$ from the angular momentum operator so its eigenvalues are just numbers.

30.2 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[3]. 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 of degree two (with real coefficients) is two dimensional. We can build a corresponding table as shown in Table 30.1.

Thus the number of edge excitations at a given angular momentum

[3]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. See footnote ****

$L - L_{ground}$	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 30.1 Table of Symmetric Polynomials

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![4])

30.2.1 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 30.2. We see the fock space is precisely equivalent to the above table of polynomials. In fact the analogy is extremely precise. In the thermodynamic limit, up to a known normalization constant, application of $b_m^\dagger$ is precisely equivalent to multiplication of the wavefunction by p_m.

These operators describe a *chiral* boson – chiral because they only have

[4] To get you started, consider filled states in a line filled up to the chemical potential. We can think of these as dots in a row. For example, let the ground state be

$$\ldots\bullet\bullet\bullet\bullet\bullet\bullet\circ\circ\circ\circ\ldots$$

where $\bullet$ means a filled single particle eigenstate and $\circ$ means empty. Now if we add one unit of (angular) momentum, we have the unique state

$$\ldots\bullet\bullet\bullet\bullet\bullet\circ\bullet\circ\circ\circ\ldots$$

adding two units can be done in two ways

$$\ldots\bullet\bullet\bullet\bullet\bullet\circ\circ\bullet\circ\circ\ldots$$

and

$$\ldots\bullet\bullet\bullet\bullet\circ\bullet\bullet\circ\circ\circ\ldots$$

thus starting the series $1, 2, 3, 5, 7 \ldots$.

$L - L_{ground}$	dimension	basis fock states	Energy
1	1	$b_1^\dagger\|0\rangle$	α
2	2	$b_2^\dagger\|0\rangle$, $b_1^\dagger b_1^\dagger\|0\rangle$	2α
3	3	$b_3^\dagger\|0\rangle$, $b_2^\dagger b_1^\dagger\|0\rangle$, $b_1^\dagger b_1^\dagger b_1^\dagger\|0\rangle$	3α
4	5	$b_4^\dagger\|0\rangle$, $b_3^\dagger b_1^\dagger\|0\rangle$, $b_2^\dagger b_1^\dagger b_1^\dagger\|0\rangle$, $b_1^\dagger b_1^\dagger b_1^\dagger b_1^\dagger\|0\rangle$	4α

Table 30.2 Fock Space for Chiral Bosons

positive angular momentum $m > 0$ not negative angular momentum.[5]

30.3 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.

[5] An *achiral* bose field on a circle requires both positive and negative angular momentum modes).

Conformal Field Theory Approach to Fractional Quantum Hall Effect

31

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 will 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.

31.1 The Chiral Boson and The Laughlin State

An interesting feature of theories in 1+1d is that they can often be decomposed (mostly[1]) cleanly into right moving and left moving pieces. So for example, if we take the simplest possible 1+1 d system, a free

[1] There may be issues with the decomposition, for example, in the case of the boson, there is a complication associated with the so-called zero-mode, which we will ignore for simplicity.

boson, we can write an achiral Lagrangian density for a field $\Phi(x,t)$ as

$$\mathcal{L} \propto (\partial_\mu \Phi)(\partial^\mu \Phi)$$

This can be decomposed into right and left moving pieces as

$$\Phi(x,t) = \phi(x-vt) + \bar{\phi}(x+vt)$$

where ϕ is right-moving and $\bar{\phi}$ is left-moving and these are two different fields. For simplicity we will set the velocity $v = 1$.

In the previous chapter we deduced that the edge theory of the Laughlin state could be described by a chiral boson Hamiltonian

$$H = \sum_{m>0} (\alpha m) a_m^\dagger a_m$$

Quantizing the boson lagrangian we find that[2]

$$\phi(x) = \sum_{m>0} \frac{i}{\sqrt{m}} e^{2\pi i m x/L} a_m^\dagger + \text{h.c.} \tag{31.1}$$

where L is the (periodic) length of the system.

We will often work in complex coordinates x and $\tau = it$, so we have we write $\Phi(z,z^*)$ where $z = x + i\tau$ and $z^* = x - i\tau$ correspond to right (holomorphic) and left-moving (antiholomorphic) coordinates.

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[3]

$$\langle \phi(z)\phi(z')\rangle = -\log(z-z')$$

Note that we think of this correlation function as a correlation in a 1+1d theory even though we are working with complex z.

From this chiral ϕ operator we construct the so-called vertex operators

$$V_\alpha(z) =: e^{i\alpha\phi(z)} :$$

where : : means normal ordering[4] A straightforward exercise (assigned as homework!) using Wick's theorem then shows that

$$\begin{aligned} \langle V_{\alpha_1}(z_1) V_{\alpha_2}(z_2) \dots V_{\alpha_N}(z_N)\rangle &= e^{-\sum_{i<j} \alpha_i \alpha_j \langle \phi(z_i)\phi(z_j)\rangle} \\ &= \prod_{i<j} (z_i - z_j)^{\alpha_i \alpha_j} \end{aligned} \tag{31.2}$$

so long as

$$\sum_i \alpha_i = 0 \tag{31.3}$$

(otherwise the correlator vanishes).

[2] We have dropped the zero mode here.

[3] Perhaps the easiest way to see this is to calculate directly from Eq. 31.1. See exercise ***. Another way to obtain this is to aim for 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.

[4] 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.

31.1.1 Writing the Laughlin Wavefunction

We then define an "electron operator" to be

$$\psi_e(z) = V_\alpha(z)$$

where we will choose

$$\alpha = \sqrt{m}$$

This then enables us to write the holomorphic part of the Laughlin wavefunction as

$$\Psi^{(m)}_{Laughlin} = \langle \psi_e(z_1)\psi_e(z_2)\dots\psi_e(z_N)\hat{Q}\rangle = \prod_{i<j}(z_i - z_j)^m$$

The index α must be chosen such that α^2 is an integer such that the wavefucntion 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 $\hat{Q} = V_{-N\alpha}$, i.e., a neutralizing charge at infinity such that Eq. 31.3 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![5].

[5] 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 handle 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.

31.1.2 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

$$\begin{aligned}\Psi_{qh}(w) &= \langle \psi_{qh}(w)\psi_e(z_1)\psi_e(z_2)\dots\psi_e(z_N)\hat{Q}\rangle \\ &= \left[\prod_i (z_i - w)^{\beta\sqrt{m}}\right]\Psi^{(m)}_{Laughlin}\end{aligned} \tag{31.4}$$

Since we must insist that the wavefunction is single valued in the z coordinates, we must choose

$$\beta = p/\sqrt{m}$$

for some positive integer p, where the minimally charged quasiparticle is then obviously $p = 1$. (Negative p is not allowed as it would create poles in the wavefunction).

Further, using this value of the the charge β, along with the smeared out background charge, we correctly obtain the normalizing gaussian

factor for the quasiparticle

$$e^{-|w|^2/(4m\ell^2)}$$

This is the correct gaussian factor, with an exponent $1/m$ times as big because the charge $V_{1/\sqrt{m}}$ is $1/m$ times as big as that of the electron charge $V_{\sqrt{m}}$.

If we are now to add multiple quasiholes, we obtain the wavefunction

$$\begin{aligned} \Psi(w_1,\dots,w_M) &= \langle \psi_{qh}(w_1)\dots\psi_{qh}(w_M)\,\psi_e(z_1)\dots\psi_e(z_N)\,Q\rangle \quad (31.5)\\ &= C\prod_{\alpha<\beta}(w_\alpha-w_\beta)^{1/m}\prod_{\alpha=1}^{M}e^{-|w_\alpha|^2/(4\ell^{*2})}\left[\prod_{\alpha=1}^{M}\prod_{i=1}^{N}(z_i-w_\alpha)\right]\Psi^{(m)}_{Laughlin} \end{aligned}$$

which is properly normalized

$$\langle\Psi(w_1,\dots w_M)|\Psi(w_1,\dots w_M)\rangle = \text{Constant}$$

and is in holomorphic gauge. As discussed previously in chapter *** with a normalized holomorphic wavefunction we can simply read off the fractional statistics as the explicit monodromy.

Note that we can consider fusion of several quasiparticles

$$V_{1/\sqrt{m}}\times V_{1/\sqrt{m}}\to V_{2/\sqrt{m}} \tag{31.6}$$

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 identity 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 (that the resulting wavefunctions will be properly normalized) it appears to hold up in many important cases.

We hope now to generalize this construction by using more complicated conformal field theories. This then generates more complicated fractional quantum Hall wavefunctions corresponding to more complcated TQFTs.

31.2 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[6]. 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:

[6] 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.

(1) There will be a finite set[7] 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.

[7] A *nonrational* CFT may have an infinite number of particle types, but these are badly behaved and do not appear to correspond to nice TQFTs.

(2) Each primary field has a **scaling dimension**[8] or **conformal weight** or **conformal spin**, which we call h_i. The scaling dimension of I is $h_I = 0$. We have see these quantities before when we discussed twists in world lines. Often we will only be interested in h modulo 1, since the twist factor is $e^{2\pi i h}$. Each primary field has descendant fields which are like derivatives of the primary and they have scaling dimensions h_i plus an integer (we will typically not need these, but for example, $\partial_z \phi_i$ has scaling dimension $h_i + 1$).

(3) Fusion relations exist for these fields, which are associative and commutative

$$\phi_i \times \phi_j = \sum_k N_{ij}^k \phi_k$$

where fusion with the identity is trivial

$$I \times \phi_j = \phi_j$$

[8] In CFT we have the powerful relation that if we make a coordinate transform $w(z)$ then any correlator of primary fields transforms as

$$\langle \phi_{i_1}(w_1) \dots \phi_{i_N}(w_N) \rangle = \left[\left(\frac{\partial w_1}{\partial z_1} \right)^{-h_{i_1}} \dots \left(\frac{\partial w_N}{\partial z_N} \right)^{-h_{i_N}} \right] \langle \phi_{i_1}(z_1) \dots \phi_{i_N}(z_N) \rangle$$

However, we will not need this relationship anywhere for our discussion!

As with TQFTs, each particle type has a unique antiparticle. We will give a clearer meaning to these fusion relations in a moment when we discuss operator product expansion.

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

$$\lim_{w\to z} \quad \langle\psi_a(w)\psi_b(z)\ \psi_c(y_1)\psi_d(y_2)\dots\psi_n(y_m)\rangle = \sum_k C_{ab}^k (w-z)^{h_k-h_a-h_b}\langle\psi_k(z)\ \psi_c(y_1)\psi_d(y_2)\dots\psi_n(y_m)\rangle$$

31.2.1 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

$$h_{\alpha+\beta} - h_\alpha - h_\beta = \frac{(\alpha+\beta)^2}{2} - \frac{\alpha^2}{2} - \frac{\beta^2}{2} = \alpha\beta$$

Note that this fusion law for the chiral boson gives more precise meaning to the fusion law we wrote in Eq. 31.6. ***(clean this up)**

31.2.2 Example: Ising CFT

The Ising CFT is actually the CFT corresponding to a 1+1 d free 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 are (exactly as in the Ising TQFT *** previously)

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

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[9]

$$\sigma(w)\sigma(z) \sim C^I_{\sigma\sigma}(w-z)^{-1/8} I + C^\psi_{\sigma\sigma}(w-z)^{3/8}\psi(z) + \dots \qquad (31.7)$$

where all terms in the ... must have branch cuts that match one of the

[9] Remember these exponents of 1/8 and 3/8 from the Ising anyon homework problems? ***

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

$$\lim_{w\to z}\langle\sigma(w)\sigma(z)\rangle \sim (w-z)^{-1/8} \tag{31.8}$$

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

$$\lim_{w\to z}\langle\sigma(w)\sigma(z)\psi(y)\rangle \sim (w-z)^{3/8} \tag{31.9}$$

Similarly one can see that fusion of two σ's in the presence of any even number of ψ fields will be similar to Eq. 31.8, whereas in the presence of any odd number of ψ fields it will be like Eq. 31.9.

Since the Ising CFT is actually a free fermion theory, we can use Wick's (fermionic) theorem for correlators of the ψ fermi fields with the added information that[10,11]

$$\langle\psi(z)\psi(w)\rangle = \frac{1}{z-w}$$

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

[10] Insert footnote or appendix that derives this. See Yellow Book for now!

[11] Add footnote on wick's theorem?***

31.3 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 and are simply multiplied together.

We then look at the operator product expansion to see what happens

when two electrons approach each other

$$\psi_e(z)\psi_e(w) \sim \left[\frac{I}{z-w}\right]\left[(z-w)^{\alpha^2}V_{2\alpha}\right]$$

where the first bracket is from the Ising part of the theory and the second bracket is from the bose part of the theory. 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

$$\psi_e(z) = \psi(z)V_{\sqrt{m}}(z)$$

with $m \geq 1$. Using this proposed electron operator we build the multiparticle 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 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

$$\begin{aligned}\langle \psi(z_1)\psi(z_2)\dots\psi(z_N)\rangle &= \mathcal{A}\left[\frac{1}{z_1-z_2}\,\frac{1}{z_3-z_4}\dots\frac{1}{z_{N-1}-z_N}\right] \\ &\equiv \mathrm{Pf}\left(\frac{1}{z_i-z_j}\right) \qquad (31.10)\end{aligned}$$

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"[12]. Thus we obtain the trial wavefunction based on the Ising CFT

$$\Psi = \mathrm{Pf}\left(\frac{1}{z_i-z_j}\right)\prod_{i<j}(z_i-z_j)^m \prod_{i=1}^{N} e^{-|z_i|^2/(4\ell^2)}$$

which is known as the Moore-Read wavefunction. For m odd this is a

[12] 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

$$\mathrm{Pf}[M] = \mathcal{A}[M_{12}M_{34}\dots].$$

Also it is useful to know that $(\mathrm{Pf}[M])^2 = \det M$.

wavefunction for bosons 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 $\nu = 1/m$.

31.3.1 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[13].

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[14] for electrons at $\nu = 5/2$ interacting with the usual Coulomb interaction. This is very suggestive that the $\nu = 5/2$ is topologically equivalent to the Moore-Read Pfaffian wavefunction (i.e., they are in the same phase of matter)[15] 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.

[13] 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

$$\lim_{z_3 \to z_1} \psi_e(z_3)V_2(z_1) \sim (z_3 - z_1)^2\psi V_3(z_1)$$

[14] Here we have used a mapping between Landau levels, that any partially filled higher Landau level can be mapped to a partially filled lowest Landau level at the price of modifying the inter-electron interaction. This mapping is exact to the extent that there is no Landau level mixing. I.e., that the spacing between Landau levels is very large.

[15] 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! The breaking of the particle-hole symmetry is very weak and involves Landau-level mixing. From numerics it appears that the $\nu = 5/2$ state is actually in the phase of matter defined by the conjugate of the Moore-Read state. *** add refs

31.4 Quasiholes of the Moore-Read state

We now try to construct quasiholes for the Moore-Read Pfaffian wavefunction. As we did in Eq. 31.4, 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 (***include fields on the right? ***)

$$\psi^L_{qh}(w)\psi_e(z) \sim (w-z)^{\beta\sqrt{m}}\psi(z)$$

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

$$\begin{aligned}\Psi^L_{qh}(w) &= \langle \psi^L_{qh}(w)\,\psi_e(z_1)\psi_e(z_2)\dots\psi_e(z_N)\hat{Q}\rangle \qquad (31.11)\\ &= \left[\prod_{i=1}^{N}(z_i - w)\right]\Psi^{(m)}_{Moore-Read}\end{aligned}$$

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

$$\psi_{qh}(w) = \sigma(w)V_\beta(w)$$

We then have the operator product expansion

$$\psi_{qh}(w)\psi_e(z) \sim [\sigma(w)\psi(z)]\left[V_\beta(w)V_{\sqrt{m}}(z)\right] \sim (w-z)^{-1/2}(w-z)^{\beta\sqrt{m}}$$

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 31.2, 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[16]. We thus consider the wavefunction of the form

$$\Psi_{qh}(w,w') = \langle \psi_{qh}(w)\psi_{qh}(w')\,\psi_e(z_1)\psi_e(z_2)\dots\psi_e(z_N)\hat{Q}\rangle \quad (31.12)$$

$$= (w-w')^{\frac{1}{4m}}\; e^{-(|w|^2+|w'|^2)/4\ell^{*2}} \prod_{i=1}^{N}(w-z_i)^{1/2}(w'-z_i)^{1/2} \quad (31.13)$$

$$\times \quad \langle \sigma(w)\sigma(w')\,\psi(z_1)\psi(z_2)\dots\psi(z_N)\rangle \;\prod_{i<j}(z_i-z_j)^m \prod_{i=1}^{N} e^{-|z_i|^2/(4\ell^2)}$$

Several comments are in order here. First of all, from the first line of Eq. 31.13 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, similar to the Pfaffian of Eq. 31.10). 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 wavefunction, the vanishing property of the wavefunction remains the same, and thus this is an exact zero energy state of the special interaction.

A Crucial Assumption

The wavefunction here is single valued in all electron coorrdinates (as it should be) and is holomorphic in all coordinates (all z's and w's) except for the gaussian exponential factors. In this holomorphic gauge, as discussed above, we can read off the fractional statistics of the quasiparticles *given the assumption that the wavefunction is properly normalized.* This is a crucial assumption and it is not a simple result of CFT, but always requires an assumption about some sort of plasma being in a screening phase — and often the mapping to a plasma is highly nontrivial[17].

[16] Like the Sith, they come in pairs.

[17] See work by Bonderson et al ***.

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. 31.8. As the two quasiholes approach each other we then have[18] (** insert also h-h-h derivation of R? **)

$$\psi_{qh}(w)\psi_{qh}(w') \sim (w-w')^{\frac{1}{4m}-\frac{1}{8}}$$

where the $\frac{1}{4m}$ is written expliclty in the first line of Eq. 31.13 and the $-\frac{1}{8}$ is from the operator product expansion Eq. 31.8. 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 ***(move I downstairs here to fit with our conventions?, change notation "I" to 2qh-I?)***

$$[R^{\text{“}I\text{”}}_{qh-qh}]^2 = e^{2\pi i(\frac{1}{4m}-\frac{1}{8})}$$

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 $h_{\text{“}I\text{”}} = 1/2m$.

[18] Strictly speaking on the right hand side we should also write the identity operator I for the Ising theory and $V_{1/\sqrt{m}}$ for the boson sector.

Fusion and Braiding of Two Quasiholes in ψ Channel (odd number of electrons)

Let us now assume that the number of electrons is odd. In this case the two σ's of the quasiholes fuse to ψ as in Eq. 31.9. As the two quasiholes approach each other we then have[19]

$$\psi_{qh}(w)\psi_{qh}(w') \sim (w-w')^{\frac{1}{4m}+\frac{3}{8}}$$

where the $\frac{1}{4m}$ is written expliclty in the first line of Eq. 31.13 and the $\frac{3}{8}$ is from the operator product expansion Eq. 31.9. 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

[19] Strictly speaking on the right hand side we should also write the operator ψ for the Ising theory and $V_{1/\sqrt{m}}$ for the boson sector.

we defined in earlier chapters we have

$$[R^{\text{“}\psi\text{”}}_{qh-qh}]^2 = e^{2\pi i(\frac{1}{4m}+\frac{3}{8})}$$

31.5 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. 31.13) 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

$$G(w_1, w_2, w_3, w_4) = \langle \sigma(w_1)\sigma(w_2)\sigma(w_3)\sigma(w_4)\rangle \tag{31.14}$$

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 31.2). 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 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. 31.14. 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[20].

[20] 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!

$$\lim_{w\to\infty}\langle\sigma(0)\sigma(z)\sigma(1)\sigma(w)\rangle = a_+G_+(z) + a_-G_-(z) \tag{31.15}$$

where

$$G_\pm = (wz(1-z))^{-1/8}\sqrt{1\pm\sqrt{1-z}} \tag{31.16}$$

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

$$\begin{aligned}
&\lim_{z\to 0} \quad G_+ \sim z^{-1/8} \qquad\qquad (\sigma(0)\sigma(z) \to I)\\
&\lim_{z\to 0} \quad G_- \sim z^{3/8} \qquad\qquad (\sigma(0)\sigma(z) \to \psi)
\end{aligned}$$

Thus (comparing to Eqs. 31.8 and 31.9) 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. 31.16 we see that we accumulate a phase of $e^{-2\pi i/8}$ from the factor of $(1-z)^{-1/8}$ outside the square-root. In addition, however, the $\sqrt{1-z}$ inside the square root comes back to minus itself when z wraps around 1, thus turning G_+ to G_- and vice versa! The effect of monodromy (taking z around 1) is then

$$\begin{pmatrix} a_+ \\ a_- \end{pmatrix} \longrightarrow e^{-2\pi i/8}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a_+ \\ a_- \end{pmatrix}$$

(This result should be somewhat familiar from the homework exercise on Ising anyons!)

We thus see that in this language, the multiple fusion channels are just different choices of which Riemann sheet we are considering, and the fact that braiding (monodromy) changes the fusion channel is simply the fact that moving coordinates around on a Riemann surface, you can move from one Riemann sheet to another!

So long as we can *assume* that the conformal blocks are orthonormal (see comment above on "crucial assumption" about normalization of wavefunctions. Orthonormality, is now adding a further assumption[21]) 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.

[21] 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. *** myabe move bonderson ref here?

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 31.16 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.

$$\begin{pmatrix} \tilde{G}_+ \\ \tilde{G}_- \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} G_+ \\ G_- \end{pmatrix}$$

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. 31.1

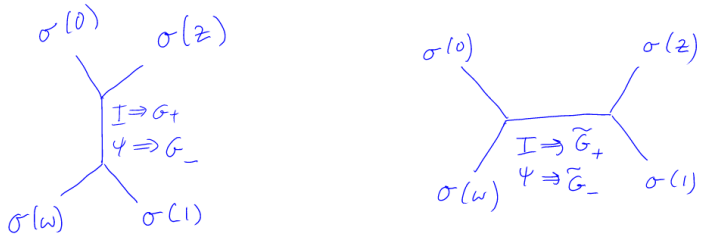


Fig. 31.1 The F-matrix transforms between the two fusion channels depicted here.

31.6 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. 31.13 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.

31.7 Generalizing to Other CFTs

The principles we used for buidling a quantum Hall state from the Ising CFT can be generalized to build quantum Hall states from other CFTs as well. The general principles are as follows:

(1) Construct an electron field which gives a ground state which is single valued in the electron coordinates. This is done bystarting with an abelian field from the CFT (one that does not have multiple fusion channels) and combining it with a chiral bose vertex operator. The filling fraction is determined entirely by the charge on the vertex operator.

(2) Identify all of the possible quasiholes by looking at all the fields in the CFT and fusing them with a chiral bose vertex operator and enforcing the condition that the electron coordinates must not have branch cuts. The charge of the quasihole is determined by the charge on the vertex operator (and the charge on the electron vertex operator).

(3) Some of the braiding properties can be determined immediately

from the operator product expansion while others require more detailed information about the form of the CFT.

31.7.1 $\mathbb{Z}_3$ 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[22]. Let us propose an electron field

$$\psi_e(z) = \psi_1(z)V_{\sqrt{m+\frac{2}{3}}}(z)$$

where m is a nonnegative integer (even for bosons, odd for fermions). It is easy to check from the OPE that

$$\psi_e(z)\psi_e(w) \sim (z-w)^m\psi_2(z)V_{2\sqrt{m+\frac{2}{3}}}(z)$$

The resulting wavefunction is then

$$\Psi = \langle\psi_e(z_1)\psi_e(z_2)\dots\psi_e(z_N)\,Q\rangle$$

which is known as the Read-Rezayi $\mathbb{Z}_3$ parafermion wavefunction.

The filling fraction of the wavefunction is determined by the vertex operator and is given by

$$\nu = \frac{1}{m+\frac{2}{3}}$$

For the $m=0$ case this is $\nu = 3/2$ bosons, while for the $m=1$ case this is $\nu = 3/5$ fermions.

For the case of $m=0$ it is easy to check that the wavefunction does not vanish when two particles come to the same point, nor does it vanish when three particles come to the same point, but it does vanish when

[22] 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 $h = 2/5 + 2/3 = 1 + 1/15$.

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. (** add cites **)

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

$$e^* = e\nu/3$$

Exercises

Exercise 31.1 *Bose Vertex Operators*

In lecture we needed the following identity

$$\langle V_{\alpha_1}(z_1)V_{\alpha_2}(z_2)\dots V_{\alpha_N}(z_N)\rangle = \prod_{i<j}(z_i - z_j)^{\alpha_i\alpha_j} \tag{31.17}$$

where

$$\sum_i \alpha_i = 0 \tag{31.18}$$

where the vertex operators are defined by

$$V_\alpha(z) =: e^{i\alpha\phi(z)} : \tag{31.19}$$

with ϕ a chiral bose field and colons meaning normal ordering.

(a) To get to this result, let us first show that for a bose operator a, such that $[a, a^\dagger] = 1$, we have

$$e^{\alpha a}e^{\beta a^\dagger} = e^{\beta a^\dagger}e^{\alpha a}e^{\alpha\beta} \tag{31.20}$$

(b) Thus derive

$$\langle V_{A_1}V_{A_2}\dots V_{A_N}\rangle = e^{\sum_{i<j}\langle A_iA_j\rangle} \tag{31.21}$$

where

$$A_i = u_i a^\dagger + v_i a \tag{31.22}$$

and

$$V_{A_i} =: e^{A_i} := e^{u_i a^\dagger}e^{v_i a} \tag{31.23}$$

with the colons meaning normal ordering (all daggers moved to the left).

(c) Show that Eq. 31.21 remains true for any operators A_i that are sums of different bose modes a_k, i.e., if

$$A_i = \sum_k [u_i(k)a_k^\dagger + v_i(k)a_k] \tag{31.24}$$

Set $A_i = i\alpha_i\phi(z_i)$ such that $V_{A_i} = V_\alpha(z_i)$. If ϕ is a free massless chiral bose field which can be written as the sum of fourier modes of bose operators such that

$$\langle\phi(z)\phi(w)\rangle = -\ln(z - w) \tag{31.25}$$

conclude that Eq. 31.17 holds.

Note: This result is not quite correct, as it fails to find the constraint Eq. 31.18 properly. The reason it fails is a subtlety which involves how one separates a bose field into two chiral components. (More detailed calculations that get this part right are given in the Big Yellow CFT book (P. Di Francesco, P. Mathieu, and D. Senechal) and in a different language in A. Tsvelik's book.)

There is, however, a quick way to see that the constraint must be true. Note that the lagrangian of a massless chiral bose field is

$$\mathcal{L} = \frac{1}{2\pi}\partial_x\phi(\partial_x + v\partial_t)\phi \tag{31.26}$$

which clearly must be invariant under the global transformation $\phi \to \phi + b$.

(d)Show that the correlator Eq. 31.17 (with Eq. 31.19) cannot be invariant under this transformation unless Eq. 31.18 is satisfied, or unless the value of the correlator is zero.

Exercise 31.2 $\mathbb{Z}_4$ Quantum Hall State

In this problem we intend to construct a quantum hall state from the the $\mathbb{Z}_4$ parafermion conformal field theory (Details of the CFT can be found in A. B. Zamolodchikov and V. A. Fateev, Soviet Physics JETP 62, 216 (1985), but we will not need too many of the details here).

The wavefunction we construct is known as the $\mathbb{Z}_4$ Read-Rezayi wavefunction (N. Read and E. Rezayi, Phys. Rev. B **59**, 8084 (1999)).

The $\mathbb{Z}_4$ parafermion conformal field theory has 10 fields with corresponding conformal weights (scaling dimension)

field	$\mathbf{1}$	ψ_1	ψ_2	ψ_3	σ_+	σ_-	ϵ	ρ	χ_+	χ_-
weight h	0	$\frac{3}{4}$	1	$\frac{3}{4}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{9}{16}$	$\frac{9}{16}$

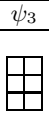
and the fusion table is given by

$\times$	$\mathbf{1}$	ψ_1	ψ_2	ψ_3	σ_+	σ_-	ϵ	ρ	χ_+	χ_-
$\mathbf{1}$	$\mathbf{1}$	ψ_1	ψ_2	ψ_3	σ_+	σ_-	ϵ	ρ	χ_+	χ_-
ψ_1	ψ_1	ψ_2	ψ_3	$\mathbf{1}$	χ_-	σ_+	ρ	ϵ	σ_-	χ_+
ψ_2	ψ_2	ψ_3	$\mathbf{1}$	ψ_1	χ_+	χ_-	ϵ	ρ	σ_+	σ_-
ψ_3	ψ_3	$\mathbf{1}$	ψ_1	ψ_2	σ_-	χ_+	ρ	ϵ	χ_-	σ_+
σ_+	σ_+	χ_-	χ_+	σ_-	$\psi_1 + \rho$	$\mathbf{1} + \epsilon$	$\sigma_+ + \chi_+$	$\sigma_- + \chi_-$	$\psi_3 + \rho$	$\psi_2 + \epsilon$
σ_-	σ_-	σ_+	χ_-	χ_+	$\mathbf{1} + \epsilon$	$\psi_3 + \rho$	$\sigma_- + \chi_-$	$\sigma_+ + \chi_+$	$\psi_2 + \epsilon$	$\psi_1 + \rho$
ϵ	ϵ	ρ	ϵ	ρ	$\sigma_+ + \chi_+$	$\sigma_- + \chi_-$	$\mathbf{1} + \psi_2 + \epsilon$	$\psi_1 + \psi_3 + \rho$	$\sigma_+ + \chi_+$	$\sigma_- + \chi_-$
ρ	ρ	ϵ	ρ	ϵ	$\sigma_- + \chi_-$	$\sigma_+ + \chi_+$	$\psi_1 + \psi_3 + \rho$	$\mathbf{1} + \psi_2 + \epsilon$	$\sigma_- + \chi_-$	$\sigma_+ + \chi_+$
χ_+	χ_+	σ_-	σ_+	χ_-	$\psi_3 + \rho$	$\psi_2 + \epsilon$	$\sigma_+ + \chi_+$	$\sigma_- + \chi_-$	$\psi_1 + \rho$	$\mathbf{1} + \epsilon$
χ_-	χ_-	χ_+	σ_-	σ_+	$\psi_2 + \epsilon$	$\psi_1 + \rho$	$\sigma_- + \chi_-$	$\sigma_+ + \chi_+$	$\mathbf{1} + \epsilon$	$\psi_3 + \rho$

If I have not made any mistake in typing this table, the fusion rules should be associative

$$(a \times b) \times c = a \times (b \times c) \tag{31.27}$$

Note of interest: These fusion rules may look mysterious, but in fact they are very closely related to the fusion rules of $SU(2)$ appropriately truncated (i.e., this is the $SU(2)_4$ WZW model). We can write each field as a young tableau with no more than 2 (for $SU(2)$) columns and no more than $4-1=3$ rows

field	$\mathbf{1}$	ψ_1	ψ_2	ψ_3	σ_+	σ_-	ϵ	ρ	χ_+	χ_-
tableau	empty									

The fusion rules are just a *slight* modification of the usual young tableau

manipulations for $SU(2)$ where columns are removed if they have 4 boxes. (See the big yellow book for details).

Using the techniques discussed in lecture:

(a) Use the operator product expansion (dimension counting) to find the singularity as two ψ_1 fields come close together. I.e, find the exponent α in the relation

$$\lim_{z' \to z} \psi_1(z')\psi_1(z) \sim (z' - z)^{\alpha}\, \psi_2(z) \tag{31.28}$$

(b) Construct all possible "electron" fields by making a product of the ψ_1 field and a chiral bose vertex operator of the form

$$\psi_e(z) = \psi_1(z) e^{i\beta\phi(z)} \tag{31.29}$$

that give a single-valued and nonsingular wavefunction for the electron. (See Eq. 31.17, but ignore the sum condition Eq. 31.18) I.e., find all acceptable values of β. Consider both the case where the "electron" is a boson or a fermion. What filling fractions do these correspond to? (There are multiple allowable solutions for both bosons and fermions). Consider among the bosonic solution, the one solution of the highest density. The ground state wavefunction in this case is the highest density zero energy state of a 5-point delta function interaction. Show that the wavefunction does not vanish when 4 particles come to the same point, but does indeed vanish as 5 particles come to the same point.

(c) Given a choice of the electron field, construct all possible quasihole operators from all fields φ in the above table

$$\phi_{qh}(w) = \varphi(w) e^{i\kappa\phi(w)} \tag{31.30}$$

For each case, fix the values of κ by insisting that the wavefunction remain single-valued in the electron coordinates. Determine the quasihole with the lowest possible (nonzero) electric charge. What is this charge?

(d) Two such quasiholes can fuse together in two possible fusion channels. What is the monodromy in each of these channels. I.e, what phase is accumulated when the two quasiholes are transported around each other (assuming the Berry matrix is zero – which is a statement about wavefunctions being properly orthonormal – which we usually assume is true).

(e) Draw a Bratteli diagram (a tree) describing the possible fusion channels for many of these elementary particles. Label the number of paths in the diagram for up to 10 quasiholes. If there are 8 quasiparticles and the number of electrons is divisible by 4, what is the degeneracy of the ground state? If there are 4 quasiparticles and the number of electrons is $4m + 2$ what is the degeneracy of the ground state?

(f) Construct a 5 by 5 transfer matrix and show how to calculate the ground state degeneracy in the presence of any number of quasiholes. Finding the largest eigenvalue of this matrix allows you to calculate the "quantum dimension" d which is the scaling

$$\text{Degeneracy} \sim d^{[\text{Number of Quasiholes}]} \tag{31.31}$$

in the limit of large number of quasiholes. While diagonalizing a 5 by 5 matrix seems horrid, this one can be solved in several easy ways (look for a trick or a nice factorization of the characteristic polynomial).

(g) Consider instead constructing a wavefunction from the ψ_2 field

$$\psi_e(z) = \psi_2(z) e^{i\beta\phi(z)} \tag{31.32}$$

What filling fraction does this correspond to (for bosons or fermions). In the highest density case, what are the properties of this wavefunction (how does it vanish as how many many electrons come to the same point).

Appendix: Resources for TQFTs

32

Working out the details of a TQFT is an often tedious task and except in the simplest cases, one does not want to go through the pain of doing this. At the end of this chapter we list a number of resources for finding details of many TQFTs.

Perhaps the most useful single resource I have found for obtaining data about TQFTs is a computer program called *Kac* written by A. N. Schellekens. The complicated part of the algorithm is described by Fuchs et al. [1996]. More details are given on the project webpage.

The progam can be downloaded from the webpage

```
https://www.nikhef.nl/~t58/Site/Kac.html
```

While the program has many capabilities (and I encourage you to RTFM[1]), it is probably useful to give here an annotated example of how it works. Note that the program uses Dynkin diagram (Cartan) notation for describing Lie algebras. The correspondence is given by

[1] Read the Frikkin Manual

$$\begin{aligned} A_r &= su(r+1) \\ B_r &= so(2r+1) \\ C_r &= sp(2r) \\ D_r &= so(2r) \end{aligned} \tag{32.1}$$

One can also use the $E6, E7, E8, F4$ and $G2$ Lie algebras.

Here we present some annotated sessions with Kac.

```
MYLINUXBOX$Kac
Kac (on MYLINUXBOX), version 8.05468, compiled on Sep  1 2016, at 16:27:29
Started Sun 30 Aug 05:19:53 BST 2020
Non-interactive mode; Assuming default answer: OK
> tensor
```

The tensor command tells the program that we might be tensoring together multiple theories.

```
> g a 1 2
```

This inputs the group (`g` for group) with the Cartan notation `a 1`, or A_1 which is $su(2)$ as given by the correspondence Eq. 32.1 above, and the `2` indicates level 2. So we are asking it to compute information about $SU(2)_2$.

```
> display
CFT {A1:2}; 3 primaries (2 simple currents)
 Lbl  Comb.  Weights     Wts.  F.l.  F.m.
   0  {0}    0.0000000   0     -     1
   1  {1}    0.5000000   1/2   -     1
   2  {2}    0.1875000   3/16  -     1
```

The fields are numbered 0,1,2, and we see their corresponding weights $h = 0, 1/2, 3/16$. The simple currents are always listed first. Recall that twist factors are given by $\theta = e^{2\pi i h}$. Note also that the weights are only correct modulo one. We can then ask for quantities like the fusion rules, or the S-matrix, the Frobenius-Schur indicator, or the central charge

```
> fusion

(0) x (0) =  (0)

(0) x (1) =  (1)

(0) x (2) =  (2)

(1) x (1) =  (0)

(1) x (2) =  (2)

(2) x (2) =  (0)  + (1)

> S

S(0,0) =      0.50000000
S(0,1) =      0.50000000
S(0,2) =      0.70710678
S(1,1) =      0.50000000
S(1,2) =     -0.70710678
S(2,2) =      0.00000000

> Get Schur 2
-1

> Browse Central
Central charge              1.500000000000000
```

If we had wanted to look at the opposite chirality theory, we use `h` rather than `g`. To wipe the memory of the program and return to tensor mode we use `reset tensor`. So for example, we have

```
> reset tensor
> h a 1 2
> display
CFT {A1:2}; 3 primaries (2 simple currents)
 Lbl  Comb.  Weights            Wts.           F.l.  F.m.
   0  {0}    0.0000000  (mod 1)  0              -     1
   1  {1}    0.5000000  (mod 1)  1/2 (mod 1)    -     1
   2  {2}    0.8125000  (mod 1)  13/16 (mod 1)  -     1
```

Note that the weight of the 2 field is $13/16 = -3/16 \mod 1$ so this is the opposite chirality version of $SU(2)_2$ which we write as $\overline{SU(2)_2}$.

The program can also handle $U(1)$ Chern-Simons theory, and accepts a parameter for so-called "radius" of the boson (which substitutes for the level of the Chern-Simons theory). Since there is some disagreement in the literature as to how you label the level of a $U(1)$ Chern-Simons theory, and as to how you label the radius, it is worth stating expliclty that in the convention used by this program, the theory with radius R has R different fields. In the convention we use in section 20.4.2 we have $U(1)_{N/2}$ corresponding to radius N for N even. We produce these theories using the code `g u` followed by the radius as follows.

```
> reset tensor
> g u 4
> display
CFT {U4:0}; 4 primaries (4 simple currents)
 Lbl  Comb.  Weights    Wts.  F.l.  F.m.
   0  {0}    0.0000000  0     -     1
   1  {1}    0.1250000  1/8   -     1
   2  {2}    0.5000000  1/2   -     1
   3  {3}    0.1250000  1/8   -     1
```

The program can handle condensation, as well as splitting. Let us consider the example used in section 23.4 of $SU(2)_4$. We first produce the $SU(2)_4$ theory

```
> reset tensor
> g a 1 4
> display
CFT {A1:4}; 5 primaries (2 simple currents)
 Lbl  Comb.  Weights    Wts.  F.l.  F.m.
   0  {0}    0.0000000  0     -     1
   1  {1}    1.0000000  1     -     1
   2  {2}    0.1250000  1/8   -     1
   3  {3}    0.6250000  5/8   -     1
   4  {4}    0.3333333  1/3   -     1
```

Note that one of the simple currents is a boson (integer weight). To condense it we issue the command `current` and the name of the field

we want to condense.

```
> current 1
> display
CFT {A1:4}; 3 primaries
 Lbl  Comb.  Weights     Wts.  F.l.  F.m.
   0  {0}    0.0000000   0     -     1
   1  {4}    0.3333333   1/3   0     1
   2  {4}    0.3333333   1/3   1     1

> fusion

(0) x (0) =  (0)

(0) x (1) =  (1)

(0) x (2) =  (2)

(1) x (1) =  (2)

(1) x (2) =  (0)

(2) x (2) =  (1)
```

Which correctly splits the 4-particle as we discussed in section 23.4.

To generate product theories, we just input several theories in a row. For example, to look at a product theory, $SU(2)_2 \times \overline{SU(2)_1} \times \overline{SU(2)_1}$ we write

```
> reset tensor
> g a 1 2
> h a 1 1
> h a 1 1
> display
CFT {A1:2_A1:1_A1:1}; 12 primaries (8 simple currents)
 Lbl  Comb.    Weights             Wts.           F.l.  F.m.
   0  {0,0,0}  0.0000000  (mod 1)  0              -     1
   1  {0,0,1}  0.7500000  (mod 1)  3/4 (mod 1)    -     1
   2  {0,1,0}  0.7500000  (mod 1)  3/4 (mod 1)    -     1
   3  {0,1,1}  0.5000000  (mod 1)  1/2 (mod 1)    -     1
   4  {1,0,0}  0.5000000  (mod 1)  1/2 (mod 1)    -     1
   5  {1,0,1}  0.2500000  (mod 1)  1/4 (mod 1)    -     1
   6  {1,1,0}  0.2500000  (mod 1)  1/4 (mod 1)    -     1
   7  {1,1,1}  0.0000000  (mod 1)  0 (mod 1)      -     1
   8  {2,0,0}  0.1875000  (mod 1)  3/16 (mod 1)   -     1
   9  {2,0,1}  0.9375000  (mod 1)  15/16 (mod 1)  -     1
  10  {2,1,0}  0.9375000  (mod 1)  15/16 (mod 1)  -     1
  11  {2,1,1}  0.6875000  (mod 1)  11/16 (mod 1)  -     1
```

Since $SU(2)_2$ has 3 fields, and each $\overline{SU(2)_1}$ has 2 fields, the product

of these three theories has 12 fields. The second column of the output shows how each field is constructed from the constituent factors. For example, the output field labeled 9 in the far left column comes from the 2 field of $SU(2)$, the 0 field from the first $\overline{SU(2)_1}$ and the 1 field from the second $\overline{SU(2)_1}$.

Let us now construct the coset $SU(2)_2/(SU(2)_1 \times SU(2)_1)$. Recall from section 23.6 that one can construct this coset by starting with $SU(2)_2 \times \overline{SU(2)_1} \times \overline{SU(2)_1}$ and condensing all possible simple current bosons. Notice in the above output that there are 8 simple currents, and the one labeled 7 or {1,1,1} is a boson. We thus issue the command

```
> current 1 1 1
> display
CFT {A1:2_A1:1_A1:1}; 3 primaries (2 simple currents)
 Lbl  Comb.    Weights              Wts.           F.l.  F.m.
   0  {0,0,0}  0.0000000  (mod 1)  0              -     1
   1  {0,1,1}  0.5000000  (mod 1)  1/2 (mod 1)    -     1
   2  {2,0,1}  0.9375000  (mod 1)  15/16 (mod 1)  -     1
```

Giving us the result that this coset is actually $\overline{\text{Ising}}$.

Further Reading

Note that many of the following references give only the so-called "modular data" for TQFTs — meaning the S-matrices (which imply the fusion rules via the Verlinde formula, Eq. 17.12) and the twist factors θ_a. However, it has recently been established that there can be cases where more than one modular TQFT can share the same modular data (Mignard and Schauenburg [2017][2]). However the simplest such case known where the modular data does not uniquely define the TQFT has 49 different particle types and for all simple TQFTs the modular data is, at least in principle, full information.

[2] See also Bonderson et al. [2019] and Delaney and Tran [2018] for discussion of what additional data might be added to make the TQFT unique

- A useful reference on conformal field theory, including WZW theories (which give you the content of the corresponding Chern-Simons theory) is given by Di Francesco et al. [1997].
- Many details of the simplest few modular tensor categories on the periodic table are given by Rowell et al. [2009]; A corresponding discussion for fermionic models is given by Bruillard et al. [2017, 2020].
- Some nice data for some simple categories is given by Bonderson [2007]. This includes, for example, the F-matrices for $SU(2)_k$ and a number of other simple theories.
- F-matrices for many more complicated theories are given by Ardonne and Slingerland [2010].
- An online database of vertex algebras and modular categories is given at

  `https://www.math.ksu.edu/~gerald/voas/`

33 Some Mathematical Basics

Many undergraduates (and even many graduates) do not get any proper education in advanced mathematics. As such I am including a very short exposition of most of what you need to know in order to read this book. For much of the book, you won't even need to know this much! If you have even a little background in mathematics you will probably know most of this already.

33.1 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 33.1 *A* **Manifold** *is a space that locally looks like a Euclidean space.*

If a manifold is bounded, contains all its limit points, and has no boundary we call it *closed.*

33.1.1 Some Simple Examples: Euclidean Spaces and Spheres

- $\mathbb{R}^n$ is obviously a manifold (it is not bounded, so therefore not closed).
- 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 the n-sphere, S^n, as points equidistant from a central point in $\mathbb{R}^{n+1}$.

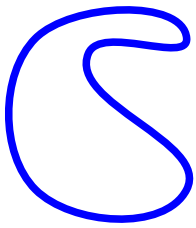


Fig. 33.1 This object is topologically a circle, S^1.

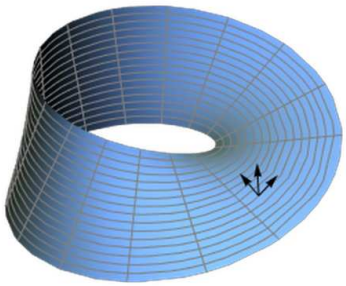


Fig. 33.2 A Möbius strip is a nonorientable manifold (with boundary). If we move the coordinate axes around the strip, when they come back to the same position, the normal vector will be pointing downwards instead of upwards.

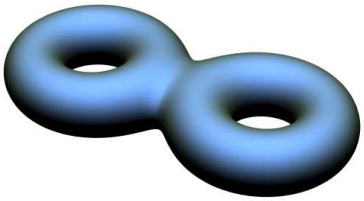


Fig. 33.3 A two handled torus is an orientable two-dimensional manifold without boundary. Because it has two holes we say it has genus two. Two dimensional manifolds without boundary are classified by their genus.

Often when we discuss a manifold, we will be interested in its topological properties only. In other words, we will not care if a circle is dented as shown in Fig. 33.1, 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 and closing up the top (which gives 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.

Orientability

We say a manifold is orientable if we can consistently define a vector normal to the manifold at all points. Another way of defining orientability (that does not rely on embedding the manifold in a higher dimension) is that we should be able to consistently define an orientation of the coordinate axes at all points on the manifold. Throughout this book we will almost always assume that all manifolds are orientable.

An example of a nonorientable manifold is the Möbius strip shown in Fig. 33.2. If we smoothy move the coordinate axes around the strip, when we come back to the same point, the upward pointing normal will have transformed into a downward facing normal.

There is a very simple classification of orientable closed (bounded and without boundary) two dimensional manifolds by the number of "holes" which is known as its "genus". A sphere has no holes, a torus has one hole, a two handled torus has two holes, and so forth. See Fig. 33.3.

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

We can take a "disjoint" union of manifolds, using the notation $\cup$. 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 (or a *disconnected manifold*). Locally it still looks like a Euclidean space.

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

One can take the product of two manifolds, or "cross" them together, using the notation $\times$. We write $\mathcal{M}_3 = \mathcal{M}_1 \times \mathcal{M}_2$. 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 2-torus T^2, or surface of a doughnut[1] 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.

[1] Alternatively spelled "donut" if you are from the states and you like coffee.

33.1.4 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. 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 some fixed radius r. Note: Often the 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 $I = D^1 = B^1$.

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[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.

[2] In the notation of Section 33.1.5 below, $\partial I = \text{pt} \cup \text{pt}$ where pt means a point and here $\cup$ means the union of the two objects as described above in 33.1.2.

- 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 $D^2 \times S^1$ $(= B^2 \times S^1)$, a 2-disc crossed with a circle.

33.1.5 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 one dimensional 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, $(\partial S^1) \times I$. 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.

33.2 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 (i.e., if $g_1 g_2 = g_2 g_1$ for all $g_1, g_2 \in G$), then we call the group **abelian**[3]. 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**[4].

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[5] called e or I 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

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

33.2.1 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 $e^{2\pi i p/N}$ 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 nonabelian. 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.
- The simplest nonabelian group is S_3. In S_3, one of the elements is

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

  Another element is

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

  where X stands for exchange (exchanges 1 and 2) and R stands

[3] Named after Abel, the Norwegian mathematician who studied such groups in the early 1800s despite living in poverty and perishing at the young age of 26 from tuberculosis. The word "abelian" is usually not capitalized due to its ubiquitous use. There are a few similar words in English which are not capitalized despite being named after people, such as "galvanic".

[4] Apparently named after someone named Nonabel.

[5] 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 or I is often most natural.

for rotate. The multiplication operation XR is meant to mean, do R first, then do X (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 R the element 1 gets moved to 2. Then when we do X the element 2 gets moved to 1. So in the product XR we have 1 getting moved back to position 1. In the end we have

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

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

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

It is easy to check that

$$X^2 = R^3 = e \tag{33.1}$$

and further we have

$$XR = R^2X \tag{33.2}$$

There are a total of 6=3! elements in the group which we can list as e, R, R^2, X, XR, XR^2. All other products can be reduced to one of these 6 posibilities using Eqs. 33.1 and 33.2.

33.2.2 More Features of Groups

A **subgroup** is a subset of elements of a group which themselves form a group. For example, the integers under addition form a group. The even integers under addition are a subgroup of the integers under addition.

The **centralizer** of an element $g \in G$ often written as $Z(g)$ is the set of all elements of the group G that commute with g. I.e., $h \in Z(g)$ iff $hg = gh$. Note that this set forms a subgroup (proof is easy!). For an abelian group G the centralizer of any element is the entire group G.

A **conjugacy class** of an element $g \in G$ is defined as the set of elements $g' \in G$ such that $g' = hgh^{-1}$ for some element $h \in G$.

Example: S_3 Above we listed some of the properties of the group S_3. S_3 has several subgroups:

- The group containing the identity element e alone
- The group containing $\{e, X\}$
- The group containing $\{e, R, R^2\}$

- The group S_3 itself (which is not a so-called "proper" subgroup)

The centralizer is just the identity element $Z(S_3) = e$, since it is the only element of the group S_3 that commutes with all elements of the group. The group has three conjugacy classes

- The identity element e
- The rotations $\{R, R^2\}$
- The reflections $\{X, XR, XR^2\}$

We can check that conjugating any element in any class gives another element within the same class. For example, consider the element X and conjugate it with the element R. We have $RXR^{-1} = XR$ which is in the same conjugacy class as X.

33.2.3 Lie Groups and Lie Algebras

A **Lie group**[6] is a group which is also a manifold. Roughly, a group with a continuous (rather than discrete) set of elements. Examples include:

[6] 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.

- 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[7]. 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, so-called, fundamental representation).
- 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.

Add something about Lie Algebra?

33.2.4 Representations of Groups:

A **representation** is a group homomorphism. This 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 an 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 $UU^\dagger = U^\dagger U = \mathbf{1}$).

A representation is reducible if the representing matrices decomposes into block diagonal form. I.e., ρ is reducible if $\rho = \rho_1 \oplus \rho_2$ for two representations ρ_1 and ρ_2. An irreducible representation is one that cannot be reduced.

An amazing fact from representation theory of discrete groups is that the number of irreducible representations of a group is equal to the number of distinct conjugacy classes.

[7] 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 $[[X, Y], Z] + [[Y, Z], X] + [[Z, X], Y] = 0$.

Irrenducible representations matrices satisfy a beautiful orthogonality relationship known as the grand orthogonality theorem (or Schur orthogonality)

$$\frac{1}{|G|}\sum_{g\in G}[\rho^R(g)]^*_{nm}[\rho^{R'}(g)]_{pq} = \delta_{np}\delta_{mq}\delta_{RR'}/d \tag{33.3}$$

where the superscript R indicates a particular representation, the subscript are the matrix elements of the ρ matrix, d is the dimension of the representation R, and $|G|$ is the total number of elements in the group.

A character is the trace of a representation matrix.

$$\chi_R(g) = \mathrm{Tr}[\rho^R(g)]$$

where the superscript R indicates we are consisdering a particular representation R. Because of the cyclic property of the trace $\mathrm{Tr}[ab] = \mathrm{Tr}[ba]$ the character is the same for all elements of a conjugacy class. One can find tables of characters for different groups in any book on group theory or on the web.

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

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

A powerful tool of topology 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 manifold. 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.

33.3.1 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.[8] Correspondingly the fundamamental group is the group with two elements $\Pi_1(SO(3)) = \mathbb{Z}_2$.

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

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).
- **Writhe and Linking Number** characterize pictures of oriented knots and links.

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, writhe, and linking, see

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

Commentary on References

34

(1) A general reference which should be useful for much of the book is the review article by Nayak, Simon, et alNayak et al. [2008].

(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 KauffmanKauffman [2001], 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 (Preskill [2004]).

(4) Frank Wilczek has two books which both discuss Berry phase and abelian anyons**?**Wilczek [1990]. 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**?**. Also consider the nice article by Richard MacKenzie**?**. 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 Witten [1989]. 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 usefulAtiyah [1988, 1997]. These are both introductions to topological quantum field theories. There is also a more detailed book by the same authorAtiyah [1990a]. The full book might be hard to read unless you have a very strong maths background.

(8) 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

(9) 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:

(10) 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

(11) I've been told the book by Jiannis Pachos on topological quantum computation is a good resource.

(12) 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

(13) 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 Sénéchal) . 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

A. Achúcarro and P. Townsend. A chern-simons action for three-dimensional anti-de sitter supergravity theories. *Physics Letters B*, 180(1):89 – 92, 1986. ISSN 0370-2693. doi: https://doi.org/10.1016/0370-2693(86)90140-1. URL `http://www.sciencedirect.com/science/article/pii/0370269386901401`.

C. C. Adams. *The knot book: an elementary introduction to the mathematical theory of knots*. W. H. Freeman and Company, 1994.

D. Aharonov and I. Arad. The bqp-hardness of approximating the jones polynomial. *New Journal of Physics*, 13(3):035019, 2011. URL `http://stacks.iop.org/1367-2630/13/i=3/a=035019`.

D. Aharonov, V. Jones, and Z. Landau. A polynomial quantum algorithm for approximating the jones polynomial. *Algorithmica*, 55, 2009. doi: https://doi.org/10.1007/s00453-008-9168-0. URL `arXiv:quant-ph/0511096`.

Y. Aharonov and D. Bohm. Significance of electromagnetic potentials in the quantum theory. *Phys. Rev.*, 115:485–491, Aug 1959. doi: 10.1103/PhysRev.115.485. URL `https://link.aps.org/doi/10.1103/PhysRev.115.485`.

R. Ainsworth and J. K. Slingerland. Topological qubit design and leakage. *New Journal of Physics*, 13(6):065030, jun 2011. doi: 10.1088/1367-2630/13/6/065030.

S. Akbulut. *4 manifolds*. Oxford Graduate Texts in Mathematics. Oxford University Press, 2016.

M. Amy. *Algorithms for the Optimization of Quantum Circuits*. PhD thesis, University of Waterloo, 2013. URL `https://uwspace.uwaterloo.ca/handle/10012/7818`.

J. F. Annette. *Superconductivity, Superfluids and Condensates*. Oxford University Press, 2004.

E. Ardonne and J. Slingerland. Clebsch–gordan and 6j-coefficients for rank 2 quantum groups. *Journal of Physics A: Mathematical and Theoretical*, 43(39):395205, aug 2010. doi: 10.1088/1751-8113/43/39/395205.

D. Arovas, J. R. Schrieffer, and F. Wilczek. Fractional statistics and the quantum Hall effect. *Phys. Rev. Lett.*, 53(7):722–3, 1984.

M. Atiyah. Topological quantum field theories. *Publications Mathématiques de l'Institut des Hautes Scientifiques*, 68, 1988. doi: https://doi.org/10.1007/BF02698547.

M. Atiyah. *The geometry and physics of knots*. 1990a.

M. Atiyah. On framings of 3-manifolds. *Topology*, 29(1):1 – 7, 1990b. ISSN 0040-9383. doi: https://doi.org/10.1016/0040-9383(90)90021-B. URL http://www.sciencedirect.com/science/article/pii/004093839090021B.

M. Atiyah. An introduction to topological quantum field theories. In *Proceedings of the 5th Gokova Geometry and Topology Conference*, Tr. J. Mathematics, page 21, 1997. URL http://www.maths.ed.ac.uk/ v1ranick/papers/atiyahinttqft.pdf.

F. Bais. Flux metamorphosis. *Nuclear Physics B*, 170(1):32 – 43, 1980. ISSN 0550-3213. doi: https://doi.org/10.1016/0550-3213(80)90474-5.

F. A. Bais and J. K. Slingerland. Condensate-induced transitions between topologically ordered phases. *Phys. Rev. B*, 79: 045316, Jan 2009. doi: 10.1103/PhysRevB.79.045316. URL https://link.aps.org/doi/10.1103/PhysRevB.79.045316.

B. Bakalov and A. Kirillov. *Lectures on Tensor Categories and Modular Functors*, volume 21 of *University Lecture Series*. American Mathematical Society, 2001.

P. Bantay. The frobenius-schur indicator in conformal field theory. *Physics Letters B*, 394(1):87 – 88, 1997. ISSN 0370-2693. doi: https://doi.org/10.1016/S0370-2693(96)01662-0. URL http://www.sciencedirect.com/science/article/pii/S0370269396016620.

J. W. Barrett and L. Crane. A lorentzian signature model for quantum general relativity. *Classical and Quantum Gravity*, 17(16): 3101–3118, aug 2000. doi: 10.1088/0264-9381/17/16/302. URL https://arxiv.org/abs/gr-qc/9904025.

J. W. Barrett and B. W. Westbury. Invariants of piecewise-linear 3-manifolds. *Transactions of the American Mathematical Society*, 348: 3997–4022, 1996. doi: https://doi.org/10.1090/S0002-9947-96-01660-1.

B. Bartlett. Fusion categories via string diagrams. *Communications in Contemporary Mathematics*, 18(05): 1550080, 2016. doi: 10.1142/S0219199715500807. URL https://doi.org/10.1142/S0219199715500807.

P. Bonderson. Splitting the topological degeneracy of non-abelian anyons. *Phys. Rev. Lett.*, 103:110403, Sep 2009. doi: 10.1103/PhysRevLett.103.110403. URL https://link.aps.org/doi/10.1103/PhysRevLett.103.110403.

P. Bonderson, M. Freedman, and C. Nayak. Measurement-only topological quantum computation. *Phys. Rev. Lett.*, 101: 010501, Jun 2008a. doi: 10.1103/PhysRevLett.101.010501. URL https://link.aps.org/doi/10.1103/PhysRevLett.101.010501.

P. Bonderson, K. Shtengel, and J. Slingerland. Interferometry of non-abelian anyons. *Annals of Physics*, 323(11):2709 – 2755, 2008b. ISSN 0003-4916. doi: https://doi.org/10.1016/j.aop.2008.01.012. URL http://www.sciencedirect.com/science/article/pii/S0003491608000171.

P. Bonderson, C. Delaney, C. Galindo, E. C. Rowell, A. Tran, and Z. Wang. On invariants of modular categories beyond modular data. *Journal of Pure and Applied Algebra*, 223(9):4065 – 4088, 2019. ISSN 0022-4049. doi: https://doi.org/10.1016/j.jpaa.2018.12.017. URL `http://www.sciencedirect.com/science/article/pii/S0022404918303098`.

P. H. Bonderson. *Non-Abelian Anyons and Interferometry*. PhD thesis, California Institute of Technology, 2007.

N. E. Bonesteel, L. Hormozi, G. Zikos, and S. H. Simon. Braid topologies for quantum computation. *Phys. Rev. Lett.*, 95(14):140503, 2005. URL `http://link.aps.org/abstract/PRL/v95/e140503`.

M. J. Bremner, C. M. Dawson, J. L. Dodd, A. Gilchrist, A. W. Harrow, D. Mortimer, M. A. Nielsen, and T. J. Osborne. Practical scheme for quantum computation with any two-qubit entangling gate. *Phys. Rev. Lett.*, 89(24):247902, Nov 2002. doi: 10.1103/PhysRevLett.89.247902.

P. Bruillard. Rank 4 premodular categories. *New York Journal of Mathematics*, 22, 2016. URL `http://nyjm.albany.edu/j/2016/22-36v.pdf`.

P. Bruillard and C. M. Ortiz-Marrero. Classification of rank 5 premodular categories. *Journal of Mathematical Physics*, 59(1):011702, 2018. doi: 10.1063/1.5020256. URL `https://doi.org/10.1063/1.5020256`.

P. Bruillard, S.-H. Ng, E. Rowell, and Z. Wang. On classification of modular categories by rank. *International Mathematics Research Notices*, 2016, 07 2015. doi: 10.1093/imrn/rnw020.

P. Bruillard, C. Galindo, T. Hagge, S.-H. Ng, J. Y. Plavnik, E. C. Rowell, and Z. Wang. Fermionic modular categories and the 16-fold way. *Journal of Mathematical Physics*, 58(4):041704, 2017. doi: 10.1063/1.4982048. URL `https://doi.org/10.1063/1.4982048`.

P. Bruillard, C. Galindo, S.-H. Ng, J. Y. Plavnik, E. C. Rowell, and Z. Wang. Classification of super-modular categories by rank. *Algebras and Representation Theory*, 23(3):795–809, Jun 2020. ISSN 1572-9079. doi: 10.1007/s10468-019-09873-9. URL `https://doi.org/10.1007/s10468-019-09873-9`.

J.-L. Brylinski and R. Brylinski. Universal quantum gates. *Mathematics of Quantum Computation, Comput. Math. Series*, page 101–116, 2002. URL `arXiv:quant-ph/0108062`.

O. Buerschaper and M. Aguado. Mapping kitaev's quantum double lattice models to levin and wen's string-net models. *Phys. Rev. B*, 80:155136, Oct 2009. doi: 10.1103/PhysRevB.80.155136. URL `https://link.aps.org/doi/10.1103/PhysRevB.80.155136`.

F. Burnell. Anyon condensation and its applications. *Annual Review of Condensed Matter Physics*, 9(1):307–327, 2018. doi: 10.1146/annurev-conmatphys-033117-054154. URL `https://doi.org/10.1146/annurev-conmatphys-033117-054154`.

F. J. Burnell, S. H. Simon, and J. K. Slingerland. Condensation of achiral simple currents in topological lattice models: Hamil-

tonian study of topological symmetry breaking. *Phys. Rev. B*, 84:125434, Sep 2011. doi: 10.1103/PhysRevB.84.125434. URL https://link.aps.org/doi/10.1103/PhysRevB.84.125434.

F. J. Burnell, S. H. Simon, and J. K. Slingerland. Phase transitions in topological lattice models via topological symmetry breaking. *New Journal of Physics*, 14(1):015004, jan 2012. doi: 10.1088/1367-2630/14/1/015004.

S. Carlip. Quantum gravity in 2 + 1 dimensions: The case of a closed universe. *Living Reviews in Relativity*, 8(1):1, Jan 2005. ISSN 1433-8351. doi: 10.12942/lrr-2005-1. URL https://doi.org/10.12942/lrr-2005-1.

R. G. Chambers. Shift of an electron interference pattern by enclosed magnetic flux. *Phys. Rev. Lett.*, 5: 3–5, Jul 1960. doi: 10.1103/PhysRevLett.5.3. URL https://link.aps.org/doi/10.1103/PhysRevLett.5.3.

Y. Chen, F. Wilczek, E. Witten, and B. Halperin. On anyon superconductivity. *Int. J. Mod. Phys. B*, 3:1001, 1989.

L. W. Clark, N. Schine, C. Baum, N. Jia, and J. Simon. Observation of laughlin states made of light. *Nature*, 582(7810):41–45, Jun 2020. ISSN 1476-4687. doi: 10.1038/s41586-020-2318-5. URL https://doi.org/10.1038/s41586-020-2318-5.

L. Crane and D. Yetter. A categorical construction of 4-d topological quantum field theories? In L. H. Kauffman and R. Baadhio, editors, *Quantum Topology*. World Scientific, Singapore, 1993.

D. Creamer. A computational approach to classifying low rank modular categories. *arXiv:1912.02269*, 2019. URL https://arxiv.org/abs/1912.02269.

S. X. Cui and Z. Wang. Universal quantum computation with metaplectic anyons. *Journal of Mathematical Physics*, 56(3):032202, 2015. doi: 10.1063/1.4914941. URL https://doi.org/10.1063/1.4914941.

C. M. Dawson and M. A. Nielsen. The solovay-kitaev algorithm. *Quantum Information and Computation*, 6:81–95, 2006. doi: https://doi.org/10.26421/QIC6.1. URL arXiv:quant-ph/0505030.

M. D. F. de Wild Propitius. *Topological Interactions in Broken Gauge Theories*. PhD thesis, Universiteit van Amsterdam, 1995. URL arxiv:hep-th/9511195v1.

C. Delaney and A. Tran. A systematic search of knot and link invariants beyond modular data. https://arxiv.org/abs/1806.02843, 2018. URL https://arxiv.org/abs/1806.02843.

P. Deligne. Catégories tensorielles: Annoyingly in french. *Moscow Mathematical Journal*, 2.2:227–248, 2002. URL http://publications.ias.edu/sites/default/files/Tensorielles.pdf.

P. Di Francesco, P. Mathieu, and D. Sénéchal. *Conformal Field Theory*. Springer, New York, 1997.

R. Dijkgraaf and E. Witten. Topological gauge theories and group cohomology. *Commun. Math. Phys.*, 129:393–429, 1990. doi: 10.1007/BF02096988.

S. Doplicher, R. Haag, and J. E. Roberts. Local observables and particle statistics. I. *Comm. Math. Phys.*, 23:199–230, 1971.

S. Doplicher, R. Haag, and J. E. Roberts. Local observables and particle statistics. II. *Comm. Math. Phys.*, 35:49–85, 1974.

W. Ehrenberg and R. E. Siday. The refractive index in electron optics and the principles of dynamics. *Proceedings of the Physical Society. Section B*, 62(1):8, 1949. URL `http://stacks.iop.org/0370-1301/62/i=1/a=303`.

I. S. Eliëns, J. C. Romers, and F. A. Bais. Diagrammatics for bose condensation in anyon theories. *Phys. Rev. B*, 90: 195130, Nov 2014. doi: 10.1103/PhysRevB.90.195130. URL `https://link.aps.org/doi/10.1103/PhysRevB.90.195130`.

P. Etingof, D. Nikshych, and V. Ostrik. On fusion categories. *Annals of Mathematics*, 162:581–642, 2005.

P. Etingof, S. Gelaki, D. Nikshych, and V. Ostrik. *Tensor Categories*. American Mathematical Society, 2015.

B. Farb and D. Margalit. *A Primer on Mapping Class Groups (PMS-49)*. Princeton University Press, 2012. ISBN 9780691147949. URL `http://www.jstor.org/stable/j.ctt7rkjw`.

R. P. Feynman and A. R. Hibbs. *Quantum Mechanics and Path Integrals*. McGraw Hill, 1965. Reprinted 2005, Dover.

B. Field and T. Simula. Introduction to topological quantum computation with non-abelian anyons. *Quantum Science and Technology*, 3(4):045004, jul 2018. doi: 10.1088/2058-9565/aacad2. URL `https://doi.org/10.1088//2F2058-9565F//aacad2`.

A. G. Fowler. Constructing arbitrary steane code single logical qubit fault-tolerant gates. *Quantum Info. Comput.*, 11(9–10):867–873, Sept. 2011. ISSN 1533-7146.

K. Fredenhagen, K. H. Rehren, and B. Schroer. Superselection sectors with braid group statistics and exchange algebras. *Commun. Math. Phys.*, 125:201–226, 1989. doi: 10.1007/BF01217906.

M. H. Freedman and Z. Wang. Large quantum fourier transforms are never exactly realized by braiding conformal blocks. *Phys. Rev. A*, 75:032322, Mar 2007. doi: 10.1103/PhysRevA.75.032322. URL `https://link.aps.org/doi/10.1103/PhysRevA.75.032322`.

M. H. Freedman, M. J. Larsen, and Z. Wang. A modular functor which is universal for quantum computation. *Commun. Math. Phys.*, 227: 605–622, 2002a. quant-ph/0001108.

M. H. Freedman, M. J. Larsen, and Z. Wang. The two-eigenvalue problem and density of Jones representation of braid groups. *Commun. Math. Phys.*, 228:177–199, 2002b.

P. Freyd, D. Yetter, J. Hoste, W. B. R. Lickorish, K. Millett, and A. Ocneanu. A new polynomial invariant of knots and links. *Bull. Amer. Math. Soc.*, 12:239–246, 1985. doi: https://doi.org/10.1090/S0273-0979-1985-15361-3.

J. Fröhlich and F. Gabbiani. Braid statistics in local quantum theory. *Rev. Math. Phys.*, 2:251–353, 1990.

J. Fuchs, A. Schellekens, and C. Schweigert. A matrix s for all simple current extensions. *Nuclear Physics B*, 473(1):323 – 366, 1996. ISSN 0550-3213. doi: https://doi.org/10.1016/0550-3213(96)00247-7. URL http://www.sciencedirect.com/science/article/pii/0550321396002477.

T. Gannon. Comments on nonunitary conformal field theories. *Nuclear Physics B*, 670(3):335 – 358, 2003. ISSN 0550-3213. doi: https://doi.org/10.1016/j.nuclphysb.2003.07.030. URL http://www.sciencedirect.com/science/article/pii/S0550321303006448.

Y. Gefen and D. J. Thouless. Detection of fractional charge and quenching of the quantum hall effect. *Phys. Rev. B*, 47: 10423–10436, Apr 1993. doi: 10.1103/PhysRevB.47.10423. URL https://link.aps.org/doi/10.1103/PhysRevB.47.10423.

D. Gepner and A. Kapustin. On the classification of fusion rings. *Physics Letters B*, 349(1):71 – 75, 1995. ISSN 0370-2693. doi: https://doi.org/10.1016/0370-2693(95)00172-H. URL http://www.sciencedirect.com/science/article/pii/037026939500172H.

P. Goddard, A. Kent, and D. Olive. Virasoro algebras and coset space models. *Physics Letters B*, 152(1):88 – 92, 1985. ISSN 0370-2693. doi: https://doi.org/10.1016/0370-2693(85)91145-1. URL http://www.sciencedirect.com/science/article/pii/0370269385911451.

R. E. Gompf and A. I. Stipsicz. *4-Manifolds and Kirby Calculus*, volume 20 of *Graduate Studies in Mathematics*. American Mathematical Society, 1999.

D. Gross, J. Eisert, N. Schuch, and D. Perez-Garcia. Measurement-based quantum computation beyond the one-way model. *Phys. Rev. A*, 76:052315, Nov 2007. doi: 10.1103/PhysRevA.76.052315. URL https://link.aps.org/doi/10.1103/PhysRevA.76.052315.

B. I. Halperin. Statistics of quasiparticles and the hierarchy of fractional quantized Hall states. *Phys. Rev. Lett.*, 52(18):1583–6, 1984.

M. Hamermesh. *Group Theory and its Application to Physical Problems*. Dover, 1989.

A. Harrow. Quantum compiling. Ph.D. Thesis. Available online at http://www.media.mit.edu/physics/publications/theses/01.05.aram.pdf, 2001.

B. Hasslacher and M. J. Perry. Spin networks are simplicial quantum gravity. *Physics Letters B*, 103(1):21 – 24, 1981. ISSN 0370-2693. doi: https://doi.org/10.1016/0370-2693(81)90185-4. URL http://www.sciencedirect.com/science/article/pii/0370269381901854.

S.-M. Hong. On symmetrization of 6j-symbols and levin-wen hamiltonian. *arXiv:0907.2204*, 2009.

L. Hormozi, G. Zikos, N. E. Bonesteel, and S. H. Simon. Topological quantum compiling. *Phys. Rev. B*, 75(16):165310, 2007. URL `http://link.aps.org/abstract/PRB/v75/e165310`.

L. Hormozi, N. E. Bonesteel, and S. H. Simon. Topological quantum computing with read-rezayi states. *Phys. Rev. Lett.*, 103: 160501, Oct 2009. doi: 10.1103/PhysRevLett.103.160501. URL `https://link.aps.org/doi/10.1103/PhysRevLett.103.160501`.

Y. Hu, Y. Wan, and Y.-S. Wu. Twisted quantum double model of topological phases in two dimensions. *Phys. Rev. B*, 87: 125114, Mar 2013. doi: 10.1103/PhysRevB.87.125114. URL `https://link.aps.org/doi/10.1103/PhysRevB.87.125114`.

C. Jones and D. Penneys. Operator algebras in rigid c*-tensor categories. *Communications in Mathematical Physics*, 355(3):1121–1188, Nov 2017. ISSN 1432-0916. doi: 10.1007/s00220-017-2964-0. URL `https://doi.org/10.1007/s00220-017-2964-0`.

V. F. R. Jones. A polynomial invariant for knots via von neumann algebras. *Bulletin of the American Mathematical Society*, 12:103–112, 1985.

L. Kauffman. State models and the Jones polynomial. *Topology*, 26: 395–407, 1987.

L. Kauffman and S. Lins. *Temperley Lieb Recoupling theory and invariants of 3-manifolds.*, volume 134 of *Ann. Math. Stud.* Princeton Univ. Press, 1994.

L. H. Kauffman. *Knots and Physics, 3ed.* World Scientific, 2001.

R. Kirby. A calculus for framed links in s^3. *Inventiones Mathematicae*, 45:35–56, 1978.

R. Kirby and P. Melvin. Canonical framings for 3-manifolds. *Turkish J. Math; Proceedings of 6th Gokova Geometry-Topology conference*, 23 (1):89–116, 1999. URL `https://arxiv.org/abs/math/9812086`.

R. C. Kirby. *The Topology of 4-Manifolds*, volume 6 of *Nankai Institute of Mathematics, Tianjin, P.R. China*. Springer, 1989.

A. Y. Kitaev. Fault-tolerant quantum computation by anyons. *Ann. Phys. (N.Y.)*, 303:2, 2003. quant-ph/9707021.

A. Y. Kitaev. Anyons in an exactly solved model and beyond. *Ann. Phys. (N.Y.)*, 321:2–111, 2006. cond-mat/0506438.

V. Kliuchnikov, A. Bocharov, and K. M. Svore. Asymptotically optimal topological quantum compiling. *Phys. Rev. Lett.*, 112: 140504, Apr 2014. doi: 10.1103/PhysRevLett.112.140504. URL `https://link.aps.org/doi/10.1103/PhysRevLett.112.140504`.

G. Kuperberg. How hard is it to approximate the jones polynomial? *Theory of Computing*, 11(6):183–219, 2015. doi: 10.4086/toc.2015.v011a006. URL `http://www.theoryofcomputing.org/articles/v011a006`.

M. Lackenby. A polynomial upper bound on reidemeister

moves. *Annals of Mathemtics*, 182:491–564, 2015. doi: https://doi.org/10.4007/annals.2015.182.2.3.

A. J. Leggett. *Quantum Liquids: Bose condensation and Cooper Pairing in Condensed-matter Systems*. Oxford University Press, 2006.

J. M. Leinaas and J. Myrheim. On the theory of identical particles. *Nuovo Cimento*, 37B:1, 1977.

C. Levaillant, B. Bauer, M. Freedman, Z. Wang, and P. Bonderson. Universal gates via fusion and measurement operations on $SU(2)_4$ anyons. *Phys. Rev. A*, 92:012301, Jul 2015. doi: 10.1103/PhysRevA.92.012301. URL `https://link.aps.org/doi/10.1103/PhysRevA.92.012301`.

M. A. Levin and X.-G. Wen. String-net condensation: A physical mechanism for topological phases. *Phys. Rev. B*, 71(4):045110, Jan 2005.

W. Lickorish. The skein method for three manifold invariants. *Journal of Knot Theory and Its Ramifications*, 02 (02):171–194, 1993. doi: 10.1142/S0218216593000118. URL `https://doi.org/10.1142/S0218216593000118`.

W. B. R. Lickorish. A representation of orientable combinatorial 3-manifolds. *Annals of Mathematics*, 76(3):531–540, 1962. ISSN 0003486X. URL `http://www.jstor.org/stable/1970373`.

C.-H. Lin and M. Levin. Generalizations and limitations of string-net models. *Phys. Rev. B*, 89:195130, May 2014. doi: 10.1103/PhysRevB.89.195130. URL `https://link.aps.org/doi/10.1103/PhysRevB.89.195130`.

M. Lorente. Spin networks in quantum gravity. *Journal of Geometry and Symmetry in Physics*, 6:85?100, 2006. URL `https://projecteuclid.org/euclid.jgsp/1495245691`.

S. MacLane. *Categories for the working mathematician*. Springer-Verlag, 1971.

M. Mignard and P. Schauenburg. Modular categories are not determined by their modular data. https://arxiv.org/abs/1708.02796, 2017. URL `https://arxiv.org/abs/1708.02796`.

C. Mochon. Anyons from nonsolvable finite groups are sufficient for universal quantum computation. *Phys. Rev. A*, 67(2):022315, Feb 2003. doi: 10.1103/PhysRevA.67.022315.

C. Mochon. Anyon computers with smaller groups. *Phys. Rev. A*, 69(3):032306, 2004. URL `http://link.aps.org/abstract/PRA/v69/e032306`.

S. moon Hong and E. Rowell. On the classification of the grothendieck rings of non-self-dual modular categories. *Journal of Algebra*, 324(5):1000 – 1015, 2010. ISSN 0021-8693. doi: https://doi.org/10.1016/j.jalgebra.2009.11.044. URL `http://www.sciencedirect.com/science/article/pii/S0021869309006784`. Computational Algebra.

G. Moore and N. Read. Nonabelions in the fractional quantum Hall effect. *Nucl. Phys. B*, 360(2-3):362–96, 1991.

G. Moore and N. Seiberg. Taming the conformal zoo. *Physics Letters B*, 220(3):422 – 430, 1989. ISSN 0370-2693. doi: https://doi.org/10.1016/0370-2693(89)90897-6. URL http://www.sciencedirect.com/science/article/pii/0370269389908976.

M. Müger. Abstract duality for symmetric tensor *-categories. In J. Butterfield and J. Earman, editors, *Handbook of the Philosophy of Physics*, pages 865–922. Kluwer Academic Press, 2007. URL https://arxiv.org/abs/math-ph/0602036.

C. Nayak, S. H. Simon, A. Stern, M. Freedman, and S. Das Sarma. Non-abelian anyons and topological quantum computation. *Rev. Mod. Phys.*, 80:1083–1159, Sep 2008. doi: 10.1103/RevModPhys.80.1083. URL https://link.aps.org/doi/10.1103/RevModPhys.80.1083.

T. Neupert, H. He, C. von Keyserlingk, G. Sierra, and B. A. Bernevig. Boson condensation in topologically ordered quantum liquids. *Phys. Rev. B*, 93:115103, Mar 2016. doi: 10.1103/PhysRevB.93.115103. URL https://link.aps.org/doi/10.1103/PhysRevB.93.115103.

H. Nicolai and K. Peeters. Loop and spin foam quantum gravity: A brief guide for beginners. In I.-O. Stamatescu and E. Seiler, editors, *Approaches to Fundamental Physics: An Assessment of Current Theoretical Ideas*, pages 151–184, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg. ISBN 978-3-540-71117-9. doi: 10.1007/978-3-540-71117-9_9. URL https://arxiv.org/abs/hep-th/0601129.

H. Nicolai, K. Peeters, and M. Zamaklar. Loop quantum gravity: an outside view. *Classical and Quantum Gravity*, 22(19): R193–R247, sep 2005. doi: 10.1088/0264-9381/22/19/r01. URL https://arxiv.org/abs/hep-th/0501114.

M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, 2000.

H. Ooguri. Topological lattice models in four dimensions. *Modern Physics Letters A*, 07(30):2799–2810, 1992. doi: 10.1142/S0217732392004171. URL https://doi.org/10.1142/S0217732392004171.

V. Pasquier. Operator content of the ADE lattice models. *Journal of Physics A: Mathematical and General*, 20(16):5707–5717, nov 1987. doi: 10.1088/0305-4470/20/16/043.

R. Penrose. Angular momentum: an approach to combinatorial spacetime. In T. Bastin, editor, *Quantum Theory and Beyond*, page 151?180. Cambridge, 1971. URL http://math.ucr.edu/home/baez/penrose/Penrose-AngularMomentum.pdf.

G. Ponzano and T. Regge. Semiclassical limit of racah coefficients. In F. Bloch, editor, *Spectroscopic and group theoretical methods in physics*, pages 1–58. North-Holland Publ. Co., Amsterdam, 1968. URL http://math.ucr.edu/home/baez/penrose/Penrose-AngularMomentum.pdf.

V. V. Prasolov and A. B. Sossinsky. *Knots, Links, Braids and 3-Manifolds: An introduction to the New Invariants in Low-*

Dimensional Topology, volume 154 of *Translations of Mathematical Monographs*. American Mathematical Society, 1996.

J. Preskill. Lecture notes for physics 219:quantum computation. available at http://www.theory.caltech.edu/∼preskill/ph219/ph219_2004.html, 2004.

R. Rajaraman. *Solitons and Instantons*. North-Holland, 1982.

R. Raussendorf and H. J. Briegel. A one-way quantum computer. *Phys. Rev. Lett.*, 86:5188–5191, May 2001. doi: 10.1103/PhysRevLett.86.5188. URL https://link.aps.org/doi/10.1103/PhysRevLett.86.5188.

T. Regge. General relativity without coordinates. *Il Nuovo Cimento (1955-1965)*, 19(3):558–571, Feb 1961. ISSN 1827-6121. doi: 10.1007/BF02733251. URL https://doi.org/10.1007/BF02733251.

T. Regge and R. M. Williams. Discrete structures in gravity. *Journal of Mathematical Physics*, 41(6):3964–3984, 2000. doi: 10.1063/1.533333. URL https://doi.org/10.1063/1.533333.

K. H. Rehren. Braid group statistics and their superselection rules. In D. Kastler, editor, *The Algebraic Theory of Superselection Sectors*. World Scientific, 1990. doi: 10.1142/1093. URL https://www.worldscientific.com/doi/abs/10.1142/1093.

N. Y. Reshetikhin and V. G. Turaev. Invariants of 3-manifolds via link polynomials and quantum groups. *Invent. Math.*, 103:547?597, 1991.

J. Roberts. Skein theory and turaev-viro invariants. *Topology*, 34(4):771 – 787, 1995. ISSN 0040-9383. doi: https://doi.org/10.1016/0040-9383(94)00053-0. URL http://www.sciencedirect.com/science/article/pii/0040938394000530.

J. Roberts. Kirby calculus in manifolds with boundary. *Turkish J. Math; Proceedings of 5th Gokova Geometry-Topology conference, 1996*, 21 (1):111–117, 1997. URL https://arxiv.org/abs/math/9812086.

D. Rolfson. *Knots and Links*. AMS Chelsea Publishing, 1976.

C. Rovelli. Notes for a brief history of quantum gravity. *arXiv:gr-qc/0006061*, 2000. URL https://arxiv.org/pdf/gr-qc/0006061.pdf.

C. Rovelli. Loop quantum gravity. *Living Reviews in Relativity*, 11 (1):5, Jul 2008. ISSN 1433-8351. doi: 10.12942/lrr-2008-5. URL https://doi.org/10.12942/lrr-2008-5.

E. Rowell, R. Strong, and Z. Wang. On classification of modular tensor categories. *Communications in Mathematical Physics*, 292, 2009.

A. Schellekens. Fixed point resolution in extended wzw models. *Nuclear Physics B*, 558(3):484 – 502, 1999. ISSN 0550-3213. doi: https://doi.org/10.1016/S0550-3213(99)00476-9. URL http://www.sciencedirect.com/science/article/pii/S0550321399004769.

S. H. Simon. Quantum computing with a twist. *Physics World*, pages 35–40, September 2010.

S. H. Simon, N. E. Bonesteel, M. H. Freedman, N. Petrovic, and L. Hormozi. Topological quantum computing with only one mobile quasiparticle. *Phys. Rev. Lett.*, 96(7):070503, 2006. URL `http://link.aps.org/abstract/PRL/v96/e070503`.

A. Sossinsky. *Knots: Mathematics with a Twist*. Harvard University Press, 2002.

A. Stoimenow. Tait's conjectures and odd crossing number amphicheiral knots. *Bull. Amer. Math. Soc.*, 45:5–291, 2008. doi: https://doi.org/10.1090/S0273-0979-08-01196-8.

D. C. Tsui, H. L. Stormer, and A. C. Gossard. Two-dimensional magnetotransport in the extreme quantum limit. *Phys. Rev. Lett.*, 48(22): 1559–62, 1982.

V. Turaev and A. Virelizier. *Monoidal Categories and Topological Field theory*, volume 322 of *Progress in Mathematics*. Birkhauser, 2017.

V. G. Turaev. Topology of shadows. 1992.

V. G. Turaev. *Quantum Invariants of Knots and 3-Manifolds*. Walter de Gruyter, Berlin, New York, 1994.

V. G. Turaev and O. Y. Viro. State sum invariants of 3-manifolds and quantum 6j-symbols. *Topology*, 31(4):865–902, 1992.

C. Vafa. Toward classification of conformal theories. *Physics Letters B*, 206(3):421 – 426, 1988. ISSN 0370-2693. doi: https://doi.org/10.1016/0370-2693(88)91603-6. URL `http://www.sciencedirect.com/science/article/pii/0370269388916036`.

S. Vandoren and P. van Nieuwenhuizen. Lectures on instantons. *arxiv/0802.1862*, 2008.

E. Verlinde. Fusion rules and modular transformations in 2d conformal field theory. *Nuclear Physics B*, 300:360 – 376, 1988. ISSN 0550-3213. doi: https://doi.org/10.1016/0550-3213(88)90603-7. URL `http://www.sciencedirect.com/science/article/pii/0550321388906037`.

K. Walker. On witten's 3-manifold invariants. available at http://canyon23.net/math/1991TQFTNotes.pdf, 1991.

A. H. Wallace. Modifications and cobounding manifolds. *Can. J. Math.*, 12:503–528, 1960.

H. Wang, Y. Li, Y. Hu, and Y. Wan. Electric-magnetic duality in the quantum double models of topological orders with gapped boundaries. *Journal of High Energy Physics*, 2020(2):30, Feb 2020. ISSN 1029-8479. doi: 10.1007/JHEP02(2020)030. URL `https://doi.org/10.1007/JHEP02(2020)030`.

Z. Wang. *Topological Quantum Computation*, volume 112 of *CBMS Regional Conference Series in Mathematics*. American Mathematical Society, New York, 2010.

F. Wilczek. Magnetic flux, angular momentum, and statistics. *Phys. Rev. Lett.*, 48(17):1144–1146, Apr 1982. doi: 10.1103/PhysRevLett.48.1144.

F. Wilczek. *Fractional Statistics and Anyon Superconductivity*. World Scientific, Singapore, 1990.

E. Witten. 2 + 1 dimensional gravity as an exactly soluble system. *Nuclear Physics B*, 311(1):46 – 78, 1988. ISSN 0550-3213. doi: https://doi.org/10.1016/0550-3213(88)90143-5. URL `http://www.sciencedirect.com/science/article/pii/0550321388901435`.

E. Witten. Quantum field theory and the Jones polynomial. *Comm. Math. Phys.*, 121:351–399, 1989.

E. Witten. Three-dimensional gravity revisited. *arXiv:0706.3359*, 2007. URL `https://arxiv.org/pdf/0706.3359.pdf`.