

Topological Quantum: Lecture Notes and Proto-Book

Steven H. Simon

©2020

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.

... and one has yet informed me that there was a typo in the previous sentence. So you are obviously not doing your job! Get with the program!

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, ... 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.

I should also mention that chapter 28 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.

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 effect¹.

A list of useful references is given etc.

¹I 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

1	Introduction: History of Topology, Knots, Peter Tait and Lord Kelvin	1
2	Kauffman Bracket Invariant and Relation to Physics	5
2.1	The idea of a knot invariant	5
2.2	Relation to Physics	8
2.2.1	Twist and Spin-Statistics	11
2.3	Bras and Kets	13
2.4	Quantum Computation with Knots	15
2.5	Some quick comments about Fractional Quantum Hall Effect	16
	Exercises	18
I	Anyons and Topological Quantum Field Theories	21
3	Particle Quantum Statistics	23
3.1	Single Particle Path Integral	23
3.2	Two Identical Particles	25
3.3	Many Identical Particles	27
3.3.1	Paths in 2+1 D, the Braid Group	28
3.3.2	Paths in 3+1 D, the Permutation Group	29
3.3.3	Building a Path Integral	30
3.4	Abelian Examples	31
3.4.1	3+1 Dimensions	31
3.4.2	2+1 Dimensions	31
3.5	Nonabelian Case	32
3.5.1	Parastatistics in 3+1 Dimensions	33
	Exercises	35
4	Aharonov-Bohm Effect and Charge-Flux Composites	39
4.1	Review of Aharonov-Bohm Effect	39
4.2	Anyons as Charge-Flux Composites	41
4.2.1	Fusion of Anyons	42
4.2.2	Anti-Anyons and the Vacuum Particle	42
4.3	Anyon Vacuum on a Torus and Quantum Memory	43
4.3.1	Quantum Memory and Higher Genus	45
4.3.2	Number of Species of Anyons	45
	Exercises	46

5	Chern-Simons Theory Basics	47
5.1	Abelian Chern-Simons Theory	47
5.2	Nonabelian Chern-Simons theory: The paradigm of TQFT	50
5.3	Appendix: Odds and Ends about Chern Simons Theory	53
5.3.1	Gauge Transforms with Nonabelian Gauge Fields	53
5.3.2	Chern Simons Action is Metric Independent	54
5.3.3	Winding Number: The Pontryagin Index	55
5.3.4	Framing of the Manifold — or doubling the theory	55
5.3.5	Chern Simons Theory as Boundary of a Four Dimensional Topological Theory	56
5.3.6	Chern Simons Canonical Quantization for the Abelian Case	56
	Exercises	57
6	Short Digression on Quantum Gravity	59
6.0.1	Why this is hard	59
6.0.2	Which Approach?	59
6.1	Some general principles?	59
6.1.1	Further Comments on Connections to Quantum Gravity	61
6.2	Appendix: No Gravity Waves in 2+1 D	62
6.3	Appendix: Relation of 2+1D GR to Chern-Simons Theory (In Brief)	63
7	Topological Quantum Field Theory	65
7.1	Paraphrasing of Atiyah's Axioms	66
7.2	Adding Particles	70
7.2.1	Particles or No-Particles	72
7.3	Building Simple 3-Manifolds	74
7.3.1	S^3 and the modular S -matrix	74
7.3.2	$S^2 \times S^1$	76
7.4	Appendix: Sewing two solid tori together	77
II	Hilbert Space and Planar Diagrammatic Algebra	79
8	Fusion and Structure of Hilbert Space	81
8.1	Basics of Particles and Fusion — The Abelian Case	81
8.2	Multiple Fusion Channels - the Nonabelian Case	82
8.2.1	Example: Fibonacci Anyons	84
8.2.2	Example: Ising Anyons	85
8.3	Fusion and the N matrices	86
8.3.1	Associativity	89
8.4	Fusion and Hilbert Space	89
	Exercises	93
9	Change of Basis and F-symbols¹	95
9.0.1	Example: Fibonacci Anyons	95

9.0.2	Pentagon	97
9.1	Appendix: F -matrix Odds and Ends	98
9.1.1	Unitarity of F	98
9.1.2	F -matrix with higher fusion multiplicities	99
9.1.3	Gauge Transforms and the F -matrix	99
	Exercises	100
10	Planar Diagrams	103
10.1	What We Mean By a Diagram	103
10.2	Diagram Rules with Physics Normalization	104
10.2.1	Causal Isotopy (vs. Full Isotopy)	106
10.2.2	Summary of Planar Diagram Rules in Physics Normalization	108
10.3	Seeking Full Isotopy Invariance: Isotopy Invariant Normalization	109
10.3.1	Futher Possible Impediments to Full Isotopy Invariance	112
10.4	Frobenius-Schur Indicator	113
10.4.1	Spin 1/2 Analogy	114
10.5	Appendix: Further Simplifying Assumptions	115
10.5.1	Pivotal Assumption	115
10.5.2	Spherical Assumption	115
10.6	Appendix: Further Details	116
10.6.1	Higher fusion multiplicities	116
10.7	$[F_a^{aaa}]_{II}$ is real	117
11	Planar Diagrams with Full 2D Isotopy	119
11.1	Diagrammatic Rules	119
11.1.1	Summary of Planar Diagram Rules For Full Isotopy Invariant Theories	122
11.2	Constraints and Examples	122
11.3	Unitarity, Negative d_a , and Frobenius-Schur, again	127
11.4	More General Diagrammatic Algebras and Important Warning	129
11.5	Appendix: Isotopy Invariant F -matrix Odds and Ends	129
11.5.1	F -matrix with higher fusion multiplicities	129
11.6	Summary of properties of F	130
12	Quantum Dimension	131
13	Some Simple Examples of Planar Diagram Algebras	133
13.1	\mathbb{Z}_2 Fusion Rules	133
13.1.1	$d = 1$ loop gas	134
13.1.2	$d = -1$ loop gas	134
13.2	Fibonacci Fusion Rules: The Branching Loop Gas	135
13.3	Ising Fusion Rules: A two species loop gas	137
13.4	Rigidity	138
	Exercises	139

14 Planar Diagrams from Groups	141
14.1 Fusion as Group Multiplication	141
14.1.1 Group Cohomology	142
14.1.2 Examples: $G = \mathbb{Z}_N$	144
14.1.3 Using Nonabelian Groups?	146
14.2 Fusion of Discrete Group Representations	147
14.2.1 F-matrices	149
14.2.2 Continuous (Lie) Group Representations?	151
Exercises	152
15 Temperley-Lieb Algebra and Jones-Kauffman Anyons	153
15.1 Jones-Wenzl Projectors	154
15.2 General Values of d	159
15.3 Unitarization	161
15.4 F-matrices	162
Exercises	163
16 State Sum TQFTs	165
16.1 Simplicial Decomposition and Pachner Moves	165
16.1.1 Two Dimensions	165
16.1.2 Three Dimensions	166
16.2 The Turaev-Viro State Sum	167
16.2.1 Proof Turaev-Viro is a Manifold Invariant	168
16.2.2 Some TQFT Properties	169
16.3 Connections to Quantum Gravity Revisited	171
16.4 Dijkgraaf-Witten Model	172
16.4.1 Other Dimensions	173
16.4.2 Further Comments	174
Exercises	175
17 Braiding and Twisting	177
17.1 Twists	177
17.2 R-matrix	177
17.3 The Hexagon	180
17.4 Ocneanu Rigidity	181
17.5 Appendix: Gauge Transforms and R	182
17.6 Appendix: Parastatistics Revisited	182
Exercises	183
18 Diagrammatic Algebra, the S-matrix and the Verlinde Relation	185
18.1 Normalizations and Loops	186
18.2 Quantum Dimensions	186
18.3 Verlinde Algebra	187
18.4 Return of Kirby Color	188
18.5 S and θ and that is all?	189
18.6 Appendix: Quantum Dimensions Satisfy the Fusion Algebra	189
18.7 Appendix: Purely Algebraic Proof of Verlinde Relation	191
Exercises	192

19 Surgery and More Complicated Manifolds	195
19.1 Simple example of surgery on a 2-manifold	196
19.2 Surgery on 3-manifolds	197
19.2.1 Lickorish-Wallace Theorem	198
19.2.2 Kirby Calculus	198
19.2.3 Witten-Reshitikhin-Turaev Invariant	199
19.3 Appendix: Details of Surgery and Meaning of Kirby Moves	201
Exercises	202
20 Quantum Error Correction and The Toric Code	205
20.1 Classical Versus Quantum Information	205
20.1.1 Memories	205
20.1.2 Errors	206
20.2 The Toric Code	207
20.2.1 Toric Code Hilbert Space	207
20.2.2 Vertex and Plaquette Operators	208
20.2.3 Building the code space	212
20.3 Errors and Error Correction	214
20.3.1 σ_x errors	214
20.3.2 σ_z errors	216
20.3.3 σ^y errors	217
20.3.4 More Comments on Errors	218
20.4 Toric Code as Topological Matter	218
20.4.1 Excitations	219
20.4.2 Braiding Properties	220
20.4.3 Modular S-matrix	222
20.4.4 Flux Binding Description	222
20.5 Robustness of the Toric Code Phase of Matter – Example of Topologically Ordered Matter	222
20.6 The Notion of Topological Order	223
21 Kitaev's Generalized Toric Code: The Quantum Double of a Group — Lattice Gauge Theory	225
21.0.1 \mathbb{Z}_N toric code	227
21.1 Ground State Degeneracy in the General Nonabelian Case	227
22 More Generalizing The Toric Code: Loop Gases and String Nets	231
22.1 Toric Code Loop Gas	231
22.1.1 Excitations of the Loop Gas	234
22.2 The Double Semion Loop Gas	236
22.2.1 Microscopic Model of Doubled Semions	237
22.2.2 Double Semion Excitations	238
22.3 General String Net	239
22.4 Doubled Fibonacci Model	241
22.4.1 Excitations	243
22.4.2 Ground State Degeneracy	244
22.5 Add details of Levin Wen Model on the lattice?	245

22.6 Appendix: S -matrix for Fibonacci Anyons	245
Exercises	246
23 Introduction to Quantum Hall — The Integer Effect	251
23.1 Classical Hall Effect	251
23.2 Two-Dimensional Electrons	251
23.3 Phenomenology of Integer Quantum Hall Effect	253
23.4 Transport in Zero Disorder	254
23.5 The Landau Problem	255
23.6 Laughlin's Quantization Argument	258
23.6.1 Byers and Yang Theorem	259
23.6.2 Quantization of Hall Conductance	259
23.7 Edge States	260
23.7.1 Landau Gauge Edge Picture for Integer Quantum Hall	261
23.8 The Halperin Refinement of Laughlin's Argument	262
Exercises	263
24 Aside: A Rapid Introduction to Topological Insulators	267
24.1 Topological Phases of Matter	267
24.1.1 Gapless Edges	268
24.2 Curvature and Chern Number	269
24.3 Symmetry Protection	270
24.4 Appendix: Chern Number is Hall Conductivity	270
25 Introduction to Fractional Quantum Hall Effect	273
25.0.1 Our Model Hamiltonian	275
25.1 Landau Level Wavefunctions in Symmetric Gauge	275
25.1.1 What We Want in a Trial Wavefunction	276
25.2 Laughlin's Ansatz	277
25.2.1 Exact statements about Laughlin Wavefunction	278
25.2.2 Real Interactions	280
25.3 Quasiparticles	280
25.3.1 Quasiholes	280
25.3.2 Quasielectrons	282
25.3.3 Fractional Charge and Statistics?	283
25.4 Digression on Berry's Phase	283
25.5 Arovas-Schrieffer-Wilczek Calculation of Fractional Statistics	284
25.6 Gauge Choice and Monodromy	286
25.6.1 Fractional Statistics Calculation: Redux	287
25.7 Appendix: Building an Effective (Chern-Simons) Field Theory	288
25.8 Appendix: Quantum Hall Hierarchy	289
Exercises	291
26 Fractional Quantum Hall Edges	293
26.1 Parabolic Confinement	293
26.2 Edges of The Laughlin State	293

26.2.1	Edge Mode Field Theory: Chiral Boson	295
26.3	Appendix: Edges and Chern-Simons theory	296
27	Conformal Field Theory Approach to Fractional Quantum Hall Effect	297
27.1	The Chiral Boson and The Laughlin State	297
27.1.1	Writing the Laughlin Wavefunction	299
27.1.2	Quasiholes	299
27.2	What We Need to Know About Conformal Field Theory	300
27.2.1	Example: Chiral Boson	302
27.2.2	Example: Ising CFT	303
27.3	Quantum Hall Wavefunction Based on Ising CFT: The Moore-Read State	304
27.3.1	Some Exact Statements About the Moore-Read Wavefunction	305
27.4	Quasiholes of the Moore-Read state	306
27.5	Multiple Fusion Channels and Conformal Blocks	309
27.6	More Comments on Moore-Read State with Many Quasiholes	312
27.7	Generalizing to Other CFTs	313
27.7.1	\mathbb{Z}_3 Parafermions (briefly)	313
	Exercises	316
28	Some Mathematical Basics	321
28.1	Manifolds	321
28.1.1	Some Simple Examples: Euclidean Spaces and Spheres	321
28.1.2	Unions of Manifolds $\mathcal{M}_1 \cup \mathcal{M}_2$	322
28.1.3	Products of Manifolds: $\mathcal{M}_3 = \mathcal{M}_1 \times \mathcal{M}_2$	322
28.1.4	Manifolds with Boundary:	323
28.1.5	Boundaries of Manifolds: $\mathcal{M}_1 = \partial\mathcal{M}_2$.	323
28.2	Groups	324
28.2.1	Some Examples of Groups	325
28.2.2	More Features of Groups	326
28.2.3	Lie Groups and Lie Algebras	327
28.2.4	Representations of Groups:	328
28.3	Fundamental Group $\Pi_1(\mathcal{M})$	329
28.3.1	Examples of Fundamental Groups	329
28.4	Isotopy, Reidemeister Moves	330
28.5	Linking and Writhe	331
	Exercises	332
29	Commentary on References	333

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 physics¹. Similarly, in the field of quantum information and quantum computation, topological ideas are extremely prominent². 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 gravity³, 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 Kelvin⁴ and his close friend Peter Tait were interested in a phenomenon of fluid flow known as a smoke ring⁵, 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 century⁶. 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

¹The 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 24 and 23 we will discuss some of the key works that this Nobel Prize honored.

²We will see this starting in chapter 20 below.

³See chapter 6.

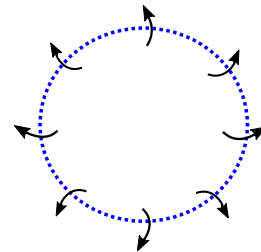


Fig. 1.1 A smoke ring or vortex loop is an invisible ring in space where the fluid flows around the invisible ring as shown by the arrows. The whole thing moves out of the plane of the page at you as the fluid circulates.

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

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

⁴Actually, 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.

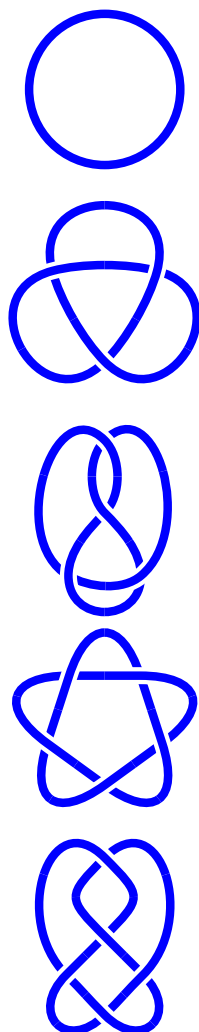


Fig. 1.2 The simplest few knots made from one strand of string. The top knot, a simple loop, is known as the “unknot”, and corresponds to the simple smoke ring in Fig. 1.1. The second knot from the top, known as the trefoil, is not the same as its mirror image (see exercise 2.1)

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 knots⁷ 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 died⁸ in 1901 not realizing that his work on the theory of knots would be important in physics, albeit for entirely different reasons.

⁷Some 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.

⁸Peter 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 cutting¹. 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.

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.

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

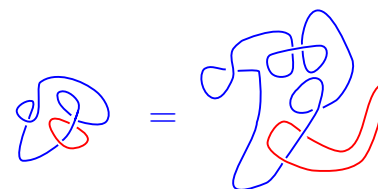


Fig. 2.1 Topological equivalence of two knots. The knot on the left can be deformed continuously into the knot on the right without cutting any strands.

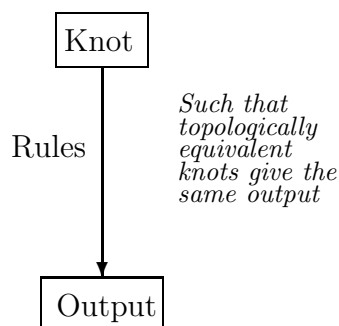


Fig. 2.2 Schematic description of a knot invariant as a set of rules taking an input knot to some mathematical output such that topologically equivalent knots give the same output.

¹A 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 28.4). This asks the question of whether there is a continuous smooth family of curves from the initial knot to the final knot — however to be more precise, as we will see below in section 2.2.1, we should think of the curves as being thickened to ribbons

²Be 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 .

³The 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 number⁴

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

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

⁵The 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.

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

$$\begin{array}{c}
 \bigcirc = -A^2 - A^{-2} = d \\
 \\
 \begin{array}{c}
 \begin{array}{c} \diagup \diagdown \\ \diagdown \diagup \end{array} = A \begin{array}{c} \curvearrowright \end{array} + A^{-1} \begin{array}{c} \curvearrowleft \end{array} \\
 \\
 \begin{array}{c} \diagdown \diagup \\ \diagup \diagdown \end{array} = A \begin{array}{c} \curvearrowleft \end{array} + A^{-1} \begin{array}{c} \curvearrowright \end{array}
 \end{array}
 \end{array}$$

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 .

$$\begin{aligned}
 & \text{Diagram of a loop with a crossing (highlighted by a red dotted circle)} = A \text{ (loop with a crossing)} + A^{-1} \text{ (loop with a crossing)} \\
 & \quad \searrow \quad \quad \quad \searrow \\
 & = A \left\{ A \text{ (loop with a crossing)} + A^{-1} \text{ (loop with a crossing)} \right\} + A^{-1} \left\{ A \text{ (loop with a crossing)} + A^{-1} \text{ (loop with a crossing)} \right\} \\
 & = A^2 d^2 + d + d^3 + A^{-2} d^2 \\
 & \quad \quad \quad \searrow \quad \quad \quad \swarrow \\
 & \quad \quad \quad -d^3 \\
 & = d
 \end{aligned}$$

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 .

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

⁷The 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.

⁸There is also some discussion of “topological” systems in 1+1 D later in section ***.

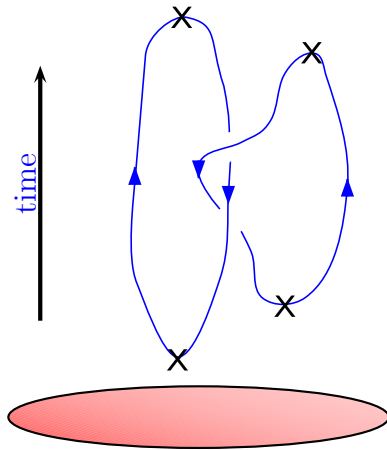


Fig. 2.5 A space-time process showing world lines of particles for a 2+1 dimensional system (shown as the shaded disk at the bottom). The X's mark the points in space-time where particles-anti-particle pairs are either pair-created or pair-annihilated.

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 times⁶ 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 cutting⁷. 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.

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

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

⁹The 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 27.

¹⁰Two Nobel Prizes have been given for work on Helium-3 superfluidity.

- | |
|--|
| <p>(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” anyons⁹. Possibly physical realizations include</p> <ul style="list-style-type: none"> • $\nu = 5/2$ Fractional Quantum Hall Effect (2D electrons at low temperature in high magnetic field). See chapters ***. • 2D p-wave superconductors. • 2D Films of ^3HeA superfluid¹⁰. • 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 ***? <p>(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.</p> <p>(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.</p> <p>(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 27). This corresponds to a fairly trivial knot invariant as we will see later in section ***.</p> <p>(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.</p> |
|--|

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.

$$\begin{aligned}
 & \text{Twisted string} = A \left(\text{Loop string} \right) + A^{-1} \left(\text{Twisted string} \right) \\
 & = \left(A[-A^2 - A^{-2}] + A^{-1} \right) = -A^3
 \end{aligned}$$

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 equivalent¹¹. 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.

¹¹In mathematics we say they are ambient isotopic but not regular isotopic! (See section 28.4)

¹²We should thus think of our knots as not just being a simple embedding of a circle S^1 into a three manifold \mathbb{R}^3 , but rather an embedding of a ribbon. This is equivalent to specifying an orthogonal vector at each point along knot which gives the orientation of the ribbon cross section at each point. When one draws an knot as a line, one must have a convention as to what this means for the orientation of the ribbon. See comment on blackboard framing at the end of this section.

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

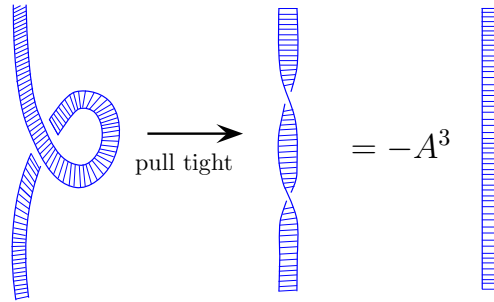


Fig. 2.7 Pulling straight a small loop introduces a twist in the string. This twist can be replaced with a factor of $-A^3$.

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 same¹⁴.

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 occurs¹⁵. This gives the famed Jones Polynomial knot invariant. See exercise 28.2.

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.

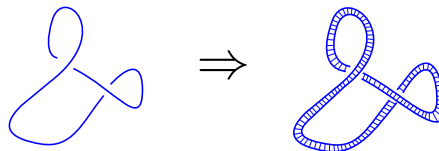


Fig. 2.8 Blackboard framing. The knot drawn on the left represents the ribbon on the right, where the ribbon always lies flat in the plane of the page (i.e., the plane of the blackboard).

¹⁴In 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.

¹⁵To properly count the self twists, one calculates the so-called “writhe” of the knot (See section 28.5). 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.

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 the ground state (or vacuum) state of some systems and create two particle-hole pairs, and imagine we tell you everything that you can locally measure about these particles (their positions, if their spin, etc etc). For most gapped systems (insulators, superconductors, charge density waves) once you know all of the locally measurable quantities, you know the full wavefunction of the system. But this is not true for topological systems. As an example, see Fig. 2.9.

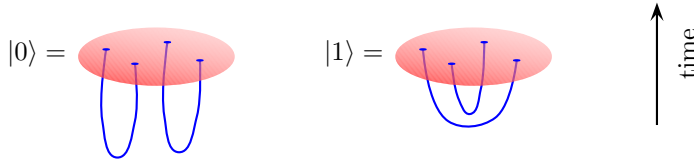


Fig. 2.9 Two linearly independent quantum states that look identical locally but have different space-time history. The horizontal plane is a space-time slice at fixed time, and the diagrams are all oriented so time runs vertically.

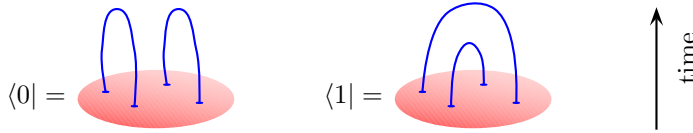


Fig. 2.10 Kets are turned into bras by reversing time.

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.

$$\begin{aligned}
\langle 0|0\rangle &= \text{[Diagram: Two vertical blue loops, each with a red oval at its base, connected by a horizontal line at the top]} = \text{[Diagram: Two separate vertical blue loops]} = d^2 \\
\langle 1|1\rangle &= \text{[Diagram: Two concentric blue loops, each with a red oval at its base]} = \text{[Diagram: Two concentric blue loops]} = d^2 \\
\langle 0|1\rangle &= \text{[Diagram: Two vertical blue loops, each with a red oval at its base, connected by a horizontal line at the top]} = \text{[Diagram: A blue figure-eight loop with a red oval at its base]} = d
\end{aligned}$$

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.

$$\begin{aligned}
\langle 0| &= \text{[Diagram: A red oval with two vertical blue loops extending upwards]} \\
&\quad \text{[Diagram: A braid between two vertical blue loops]} \\
&\quad = \langle 0|\text{Braid}|0\rangle \\
|0\rangle &= \text{[Diagram: A red oval with two vertical blue loops extending downwards]}
\end{aligned}$$

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 qubits¹⁶, and indeed this is one very good reason to be interested.

It turns out that many topological quantum systems can *compute* quantities 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 invariant¹⁷. 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 normalization¹⁸) 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 invariant 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**. One can use this so-called **topological quantum computer** to run algorithms such as Shor’s famous factoring (i.e., code breaking) algorithm²⁰. The idea of using topological systems for quantum computation is due to Michael Freedman and Alexei Kitaev²¹.

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

¹⁶One 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 ***

¹⁸If 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.

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

²⁰See Nielsen and Chuang [2000], for example, for more detail about quantum computation in general.

²¹Freedman is another Fields medalist, for his work on the Poincaré 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.

¹⁷Perhaps 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 starting thinking along the same lines and asked the question of whether one quantum system can compute the dynamics of another quantum system — which starts getting close to the ideas of modern quantum computation.

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.)

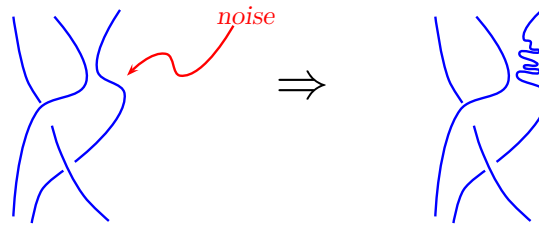


Fig. 2.13 The effect of noise on a topological quantum computation. As long as the noise does not change the topology of the knot, then no error is introduced.

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 systems²² 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 topological²³.

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

²²Electronic systems can be made two dimensional in several ways. See comments in chapter ??.

²³A 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.

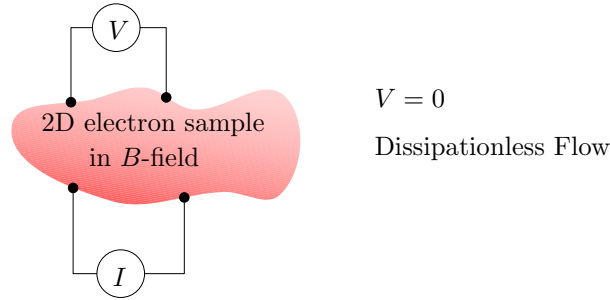


Fig. 2.14 Measurement of longitudinal resistance in FQHE experiment.

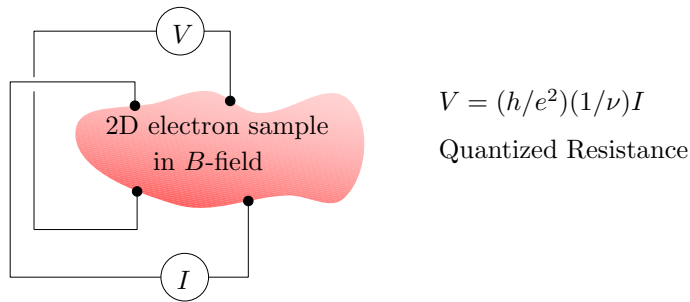


Fig. 2.15 Measurement of Hall resistance in FQHE experiment.

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!

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 (Jones) Invariant

Using the Kauffman rules, calculate the Kauffman bracket invariant of the right and left handed trefoil knots. Conclude these two knots are topologically inequivalent. While this statement appears obvious on sight, it was not proved mathematically until 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.

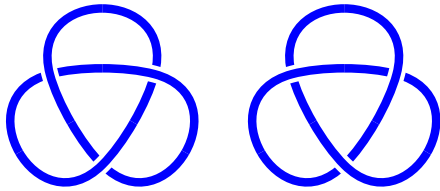


Fig. 2.16 Left and Right Handed Trefoil Knots (on the left and right respectively)

$$\text{Crossing} = e^{i\theta} \left(\text{Right Arc} \right) \left(\text{Left Arc} \right)$$

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

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

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

$$\text{Two Arcs} = \pm \left(\text{Right Arc} \right) \left(\text{Left Arc} \right)$$

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 Add an exercise on HOMFLY?

Part I

Anyons and Topological Quantum Field Theories

Particle Quantum Statistics

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

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.

¹Bose 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.

²Based 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.

³If 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.

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

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

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

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

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

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

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 first⁵. For simplicity, we can assume that $\mathbf{x}_1 \neq \mathbf{x}_2$, i.e., the particles have hard cores and cannot overlap⁶. 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 28.2)

$$\begin{array}{llll}
 \text{TYPE 1} & \text{Followed by} & \text{TYPE 1} & = \text{TYPE 1} \\
 \text{TYPE 1} & \text{Followed by} & \text{TYPE -1} & = \text{TYPE -1} \\
 \text{TYPE -1} & \text{Followed by} & \text{TYPE 1} & = \text{TYPE -1} \\
 \text{TYPE -1} & \text{Followed by} & \text{TYPE -1} & = \text{TYPE 1}
 \end{array} \quad (3.3)$$

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

Now let us try to construct a path integral, or sum over all possible

⁴Often 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 two states that have no overlap.

⁵This 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.

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

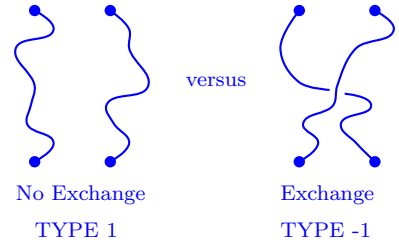


Fig. 3.1 Two possible sets of paths (paths in configuration space) from the same two starting positions to the same two ending positions (we are implying that time runs vertically). We call the non-exchange path TYPE 1, and the exchange path TYPE -1. Here we mean that time runs vertically. The two sets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

paths. It is useful to think about breaking up the sum over paths into separate sums over the two different classes of paths.

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

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

$$\begin{aligned} \langle \mathbf{x}_{1f} \mathbf{x}_{2f} | \hat{U}(t_f, t_i) | \mathbf{x}_{1i} \mathbf{x}_{2i} \rangle &= \\ \mathcal{N} \left(\sum_{\substack{\text{TYPE 1 paths} \\ i \rightarrow f}} e^{iS[\text{path}]/\hbar} - \sum_{\substack{\text{TYPE -1 paths} \\ i \rightarrow f}} e^{iS[\text{path}]/\hbar} \right) \end{aligned} \quad (3.4)$$

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

$$\begin{aligned} \langle \mathbf{x}_{1f} \mathbf{x}_{2f} | \hat{U}(t_f, t_i) | \mathbf{x}_{1i} \mathbf{x}_{2i} \rangle &= \\ \int d\mathbf{x}_{1m} d\mathbf{x}_{2m} \langle \mathbf{x}_{1f} \mathbf{x}_{2f} | \hat{U}(t_f, t_m) | \mathbf{x}_{1m} \mathbf{x}_{2m} \rangle \langle \mathbf{x}_{1m} \mathbf{x}_{2m} | \hat{U}(t_m, t_i) | \mathbf{x}_{1i} \mathbf{x}_{2i} \rangle \\ \sim \int d\mathbf{x}_{1m} d\mathbf{x}_{2m} \left(\sum_{\substack{\text{TYPE 1} \\ m \rightarrow f}} - \sum_{\substack{\text{TYPE -1} \\ m \rightarrow f}} \right) \left(\sum_{\substack{\text{TYPE 1} \\ i \rightarrow m}} - \sum_{\substack{\text{TYPE -1} \\ i \rightarrow m}} \right) e^{iS[\text{path}]/\hbar} \end{aligned}$$

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 $i \rightarrow m$ with those from $m \rightarrow f$ 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.

Our configuration space for the set of N identical particles in D dimensions can then be written as⁷.

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

Here \mathbb{R}^{ND} is a set of N coordinates in D dimensions, Δ is the space of “coincidences” where more than one particle occupy the same position (we are eliminating this possibility for simplicity). Here S_N is the group of permutations, and we are “modding” out by this group action. We said a bit about the permutation group in the mathematical section (28.2.1) on group theory, but this modding out by S_N is just a fancy way to say that we specify N coordinates, but we do not order these points (or as described above, we choose some convention for the order, like always writing the left-most first). In the case of 2 particles above, this reduced the Hilbert space by a factor of 2. More generally this should reduce the Hilbert space by a factor of $N!$. This is the same indistinguishability factor which is familiar from the Gibbs paradox of statistical mechanics.

We would now like to consider all possible paths through this configuration space \mathcal{C} . In other words we want to consider how these N different points move in time. We can think of this as a set of coordinates moving through time $\{\mathbf{x}_1(t), \dots, \mathbf{x}_N(t)\}$ but we must be careful that the particles are indistinguishable, so the order in which we write the coordinates doesn’t matter. We can think of this as N directed curves moving in $ND + 1$ dimensional space⁸. Since we want to add up all of these possible paths in a path integral it is useful to try to better understand what the structure is of this space of paths.

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. To the mathematician we are looking at the group of paths through \mathcal{C} , known as the first homotopy group $\Pi_1(\mathcal{C})$ or fundamental group⁹ (See section 28.3). The reason this is a group is that it comes with a natural operation, or

⁷This might be unfamiliar notation since S_N is not a subgroup of $(\mathbb{R}^{ND} - \Delta)$, but rather acts on $(\mathbb{R}^{ND} - \Delta)$ by permuting coordinates.

⁸The 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 annihilation, which we do not yet consider.

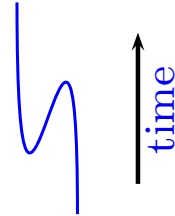


Fig. 3.2 A double-back in time is not allowed in our considerations here (and not allowed in the braid group) as it corresponds to creation and annihilation of particles at the turning around points.

⁹In 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.

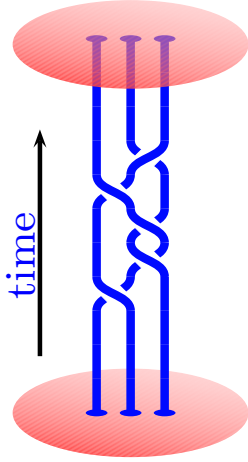


Fig. 3.3 A path through configuration space for 3 Particles in 2 dimensions (i.e, world lines in 2+1 D) is a braid with three strands.

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 28.2): 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

$$\begin{array}{lll} \sigma_1 = \begin{array}{|c|c|c|} \hline \text{X} \\ \hline \end{array} & \sigma_2 = \begin{array}{|c|c|c|} \hline \text{X} \\ \hline \end{array} & \sigma_3 = \begin{array}{|c|c|c|} \hline \text{X} \\ \hline \end{array} \\ \sigma_1^{-1} = \begin{array}{|c|c|c|} \hline \text{X} \\ \hline \end{array} & \sigma_2^{-1} = \begin{array}{|c|c|c|} \hline \text{X} \\ \hline \end{array} & \sigma_3^{-1} = \begin{array}{|c|c|c|} \hline \text{X} \\ \hline \end{array} \end{array}$$

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).

¹⁰The 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 $\sigma_i \sigma_i^{-1} = 1$.

can be written as a product of the braid generators and their inverses¹⁰. 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

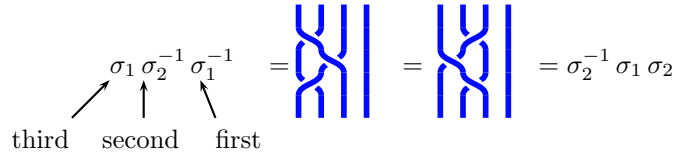


Fig. 3.5 Two braid words in B_4 that represent the same braid. The figure on the left can be continuously deformed to the one on the right, keeping endpoints fixed. The braidwords are read from right to left indicating stacking the generators from bottom to top. (The observant reader will see the similarity here to Reidemeister moves of type-III discussed in section 28.4. Similarly $\sigma_i \sigma_i^{-1} = 1$ is a type-II move.)

braid which do not change under deformation of the braid — so long as 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,¹¹ 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 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

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: But 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.

¹¹It would be very convenient to be able to draw a diagram in four dimensions!

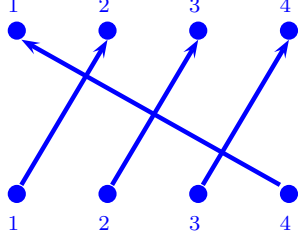


Fig. 3.7 Paths in 3+1 D are elements of the permutation group (or symmetric group) S_N (See section 28.2.1). Shown here is an element of S_4 .

S_N (see section 28.2.1). 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!

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

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

¹²In 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 $n = 1 \dots M$. 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 introduced¹² a factor of $\rho(g)$ out front where ρ is a *unitary representation* of the group G . (See section 28.2.4 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 .

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

So we have constructed all possible paths from i to f and split them into class g_2 in the region i to m and then class g_1 in the region m to f . When we compose these paths we will get a path of type $g_1 g_2$. 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 28.2.4). 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.¹³

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.

¹³We 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*¹⁴ one-dimensional representations of S_N :

¹⁴See exercise 3.2. This is a fairly short proof!

- **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 .

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**.

¹⁵There 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 Gosdard [1982], see chapter ***), since the original theoretical work was entirely abstract, and they were not thinking about any particular experiment.

¹⁶Among other things, Wilczek coined the term *anyon*. (He also won a Nobel Prize for asymptotic freedom.)

¹⁷If 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.

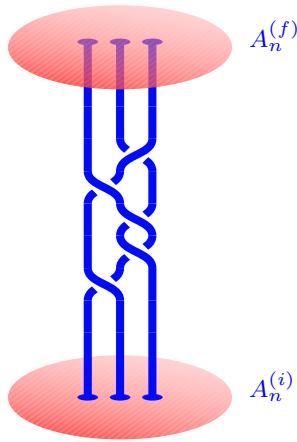


Fig. 3.8 An initial state is described by a vector $A_n^{(i)}$ multiplying the basis states $|n; \{\mathbf{x}\}_i\rangle$ as in Eq. 3.6. The particles are braided around each other in a braid g and brought back to the same positions. The final state is again described again in terms of the same basis 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.

- **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 pointed out by Leinaas and Myrheim [1977]¹⁵, and popularized by Wilczek [1982]¹⁶. 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 $n = 1 \dots M$. An arbitrary wavefunction of the system can then be expressed as

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

with the A_n 's being some complex coefficients.¹⁷ Given the N positions $\{\mathbf{x}\}$, a general wavefunction should be thought of as a vector in M dimensional complex space. Now that we have a vector, we can use an M -dimensional representation of the braid group in our path integral! We thus identify that $\rho(g)$ in Eq. 3.5 is an M by M unitary matrix

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

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

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

A particle that obeys this type of braiding statistics is known as a **non-abelian anyon**, or **nonabelion**.¹⁸ The word “nonabelian” means non-commutative, and the term is used since generically matrices (in this case the U matrices) don’t commute.

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

$$M \sim d^N \quad (3.8)$$

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 chapter 12 that (up to a possible sign) this quantum dimension d is actually the value d of the unknot¹⁹.

Some Quick Comments on Quantum Computing:

Quantum Computing is nothing more than the controlled application of unitary operations to a Hilbert space²⁰. 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.²¹

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. 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 proven²².

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

¹⁸The 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.

¹⁹Because of the possible sign, we distinguish the two quantities by using a different typeface

²⁰And initialization and measurement.

²¹The 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.

²²A 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!

each other, but also creating and annihilating. This structure is given by category theory, some parts of which we will encounter later.

- (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 17.6 where we will give further discussion of the issue.

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:

$$\sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1} \quad 1 \leq i \leq M-2 \quad (3.9)$$

$$\sigma_i \sigma_j = \sigma_j \sigma_i \quad \text{for } |i-j| > 1, \quad 1 \leq i, j \leq M-1 \quad (3.10)$$

(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 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

$$\sigma_1 \sigma_2 \dots \sigma_{M-2} \sigma_{M-1} \sigma_{M-1} \sigma_{M-2} \dots \sigma_2 \sigma_1 = I \quad (3.11)$$

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.

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

$$\{\gamma_i, \gamma_j\} \equiv \gamma_i \gamma_j + \gamma_j \gamma_i = 2\delta_{ij} \quad (3.12)$$

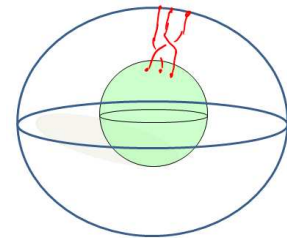


Fig. 3.9 An element of the braid group $B_3(S^2)$. The braid shown here is $\sigma_1 \sigma_2^{-1}$

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

$$U_{ij} = \frac{e^{i\alpha}}{\sqrt{2}} [1 + \gamma_i \gamma_j] \quad i < j. \quad (3.13)$$

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

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

(the notation γ^{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.

Aharonov-Bohm Effect and Charge-Flux Composites

4

This chapter introduces a simple model of how fractional statistics anyons can arise. After reviewing Aharonov-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 Aharonov-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

$$\begin{aligned} \sum_{\text{paths}} e^{iS/\hbar} &= \sum_{\text{paths, slit 1}} e^{iS/\hbar} + \sum_{\text{paths, slit 2}} e^{iS/\hbar} \\ &\sim e^{ikL_1} + e^{ikL_2} \end{aligned}$$

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 Aharonov-Bohm effect, was predicted by Ehrenberg and Siday [1949], then re-predicted independently by Aharonov and Bohm [1959]¹.

¹Possibly 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.

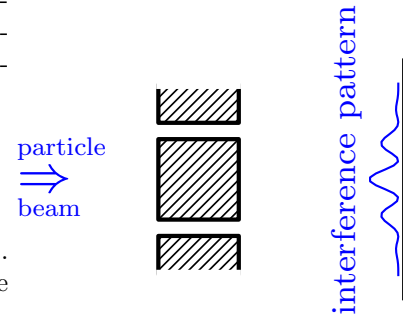


Fig. 4.1 The Young two slit experiment (not to scale).

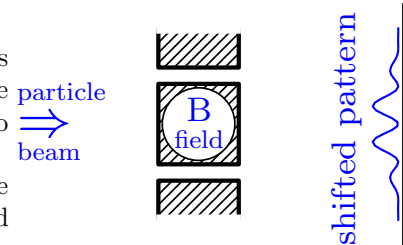


Fig. 4.2 Adding a magnetic field inside the middle box in the Young two slit experiment. Here the circular region includes a constant magnetic field. No magnetic field leaks out of the box. Nonetheless, if the particle being sent into the interferometer is charged, the interference pattern is changed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

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

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

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

$$S \rightarrow S_0 + q \int dt \dot{\mathbf{x}} \cdot \mathbf{A} = S_0 + q \int d\mathbf{l} \cdot \mathbf{A} \quad (4.2)$$

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

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

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

$$\exp \left[\frac{iq}{\hbar} \int_{\text{slit 1}} d\mathbf{l} \cdot \mathbf{A} - \frac{iq}{\hbar} \int_{\text{slit 2}} d\mathbf{l} \cdot \mathbf{A} \right] = \exp \left[\frac{iq}{\hbar} \oint d\mathbf{l} \cdot \mathbf{A} \right] \quad (4.3)$$

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

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

where Φ_{enclosed} is the flux enclosed in the loop. Thus there is a measurable relative phase shift between the two paths given by $\frac{iq}{\hbar} \Phi_{\text{enclosed}}$.

²Here 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

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

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 $\varphi = (q_1\Phi_2 + q_2\Phi_1)/\hbar$.

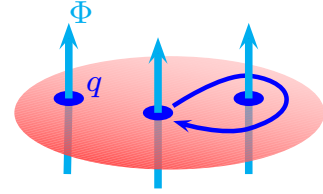


Fig. 4.3 Abelian anyons represented as charges bound to flux tubes through the plane. The charge of each particle is q , the flux of each tube is Φ . Dragging one particle around another incurs a phase both because charge is moving around a flux, but also because flux is moving around a charge.

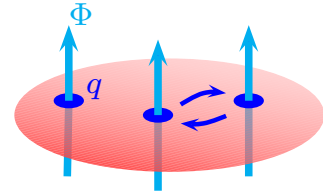


Fig. 4.4 An exchange. Two exchanges is the same as dragging one particle all the way around the other as shown in Fig. 4.3.

³Almost 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!

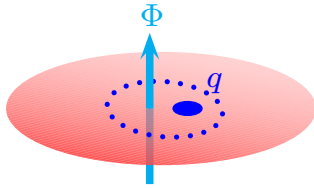


Fig. 4.5 Tying flux to charge. We put the flux and the charge at slightly different positions. As a result, when we rotate the particle around its own axis a phase is accumulated as the charge and flux go around each other.

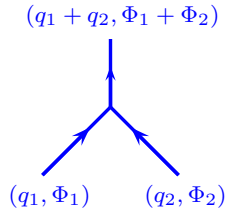


Fig. 4.6 Fusing two anyons to get an anyon of a different type which has the sum of fluxes and the sum of charges.

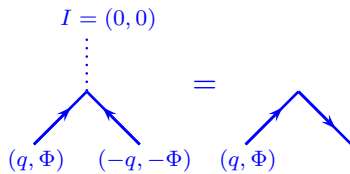


Fig. 4.7 Fusing an anyon and an anti-anyon to get the vacuum (I) drawn as dotted line. Note that the anti-anyon moving forward in time is drawn as a downpointing arrow — which looks like an anyon moving backwards in time.

⁴The 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 28.2.

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. 4.5³. 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 sometimes⁴ 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 $-\theta$.

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 doughnut⁵).

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.

Both of these operators are unitary because they can be implemented (in principle) with some time-dependent Hamiltonian⁶. 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 have⁷

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

But both T_1 and T_2 commute with the Hamiltonian (since they start and end with states of exactly the same energy⁸). 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 application of T_1 . Now we will generate a new eigenstate with a different

⁵See note 1 in chapter 28.

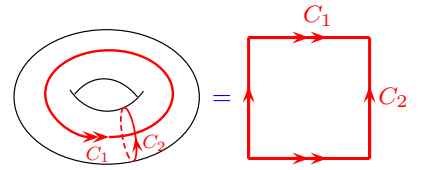


Fig. 4.8 Drawing a torus as a rectangle with opposite edges identified. The two noncontractable cycles around the torus can be considered to be the edges of the square, labeled C_1 and C_2 here.

⁶For 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.

⁷At 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 particles we create with those already present, which will be more complicated.

⁸Strictly 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.

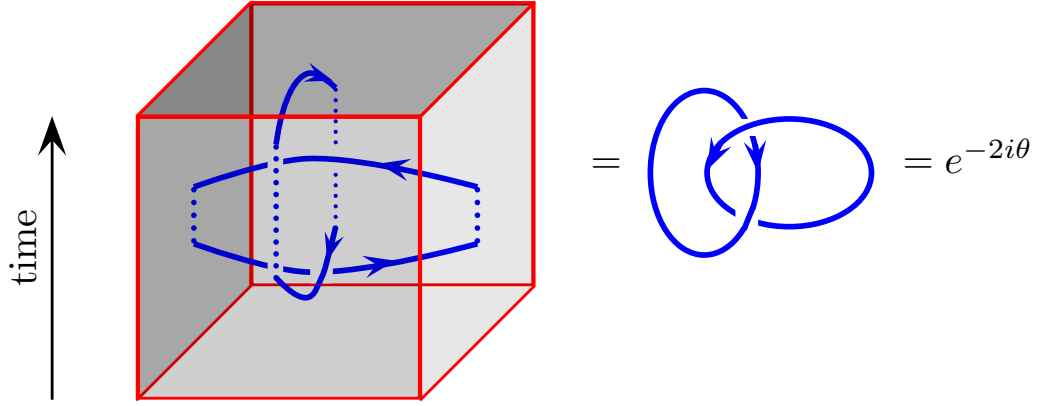


Fig. 4.9 The torus is drawn as a horizontal rectangle with opposite ends identified. Time runs vertically. First create a particle-antiparticle pair at the center of the rectangle and move them in opposite directions, right and left, until they meet at the edges of the rectangle to reannihilate. Note 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.

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 ,

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

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

⁹There 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.

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

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

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 large¹⁰. 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.

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 anyons¹² with a charge-flux composite $(q, \Phi) = (\pi p/m, 1)$. Fusion of n of these elementary anyons will have charge and flux given by¹³

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

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

$$|“m”\rangle = (\pi p, m)$$

If we braid an arbitrary cluster $|“n”\rangle = (n\pi p/m, n)$ around one of these $|“m”\rangle = (\pi p, m)$ clusters, we obtain a net phase¹⁴ of $2n\pi p$ which is equivalent to no phase at all! Thus we conclude that the cluster of

¹⁰Strictly 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.

¹¹In 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].

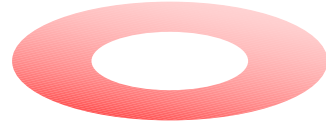


Fig. 4.10 An annulus.

¹²By this time I’m sick of writing \hbar and I’m going to set it equal to 1.

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

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

m elementary anyons is equivalent to the vacuum in the sense that all particles get trivial phase if they braid all the way around $|“m”\rangle$.

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 $|“m”\rangle = (\pi p, m)$ particle with another $|“m”\rangle = (\pi p, m)$ particle. To see this note that exchange gives a phase $\pi p m$ since it is half of the $2\pi p m$ phase for wrapping one particle all the way around the other (as in Fig. 4.3). This means the $|“m”\rangle$ particle is a fermion. In fact, this case of p and m both odd is a bit of an anomalous case and in some sense is a poorly behaved theory¹⁵.

¹⁵Whenever we have a fermion particle that braids trivially with all other particles, the theory is poorly behaved. Later on we will call this kind of theory “non-modular.” See section ***.

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

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

5.1 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 theory¹, 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

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

where j^α is the particle current, μ is some coupling constant, and ϵ is the antisymmetric tensor². 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 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

¹S. 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.

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

define the particle current j^α . If we have N particles then the current is

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

³Again 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 density³, 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

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

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

$$a_\mu \rightarrow a_\mu + \partial_\mu \chi \quad (5.3)$$

for any function χ . Then integrating 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

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

which generates the equations of motion⁴

$$j^\alpha = \mu \epsilon^{\alpha\beta\gamma} \partial_\beta a_\gamma \quad (5.5)$$

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

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

⁴It 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.

where we have defined a “Chern-Simons” magnetic field b to be the curl of 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 into⁵ 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

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

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 mean⁶ 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

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

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

⁵One 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.

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

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 considering⁷.

If we consider a three dimensional manifold of the form $\mathcal{M} = \Sigma \times S^1$ for a 2D manifold Σ and S^1 represents time (compactified⁸) 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.

⁷Some 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.

⁸Compactification 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.

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

¹⁰See the introduction to Lie groups and Lie algebras in section 28.2.3. In brief: A Lie group is a group which is also a continuous manifold, for example. 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 $g = e^{i\boldsymbol{\sigma} \cdot \mathbf{n}}$.

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-Simons 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 algebra¹⁰. 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 $i\sigma_x/2, i\sigma_y/2, i\sigma_z$ so that our Lie algebra valued gauge field is then¹¹

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

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 operators¹²

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

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 analogue¹³ 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 ordered integral is then to divide the path into tiny pieces of length dl . We then have

$$P \exp \left(\oint_L dl^\mu a_\mu \right) = [1 + a_\mu(x_1)dl^\mu(x_1)] [1 + a_\mu(x_2)dl^\mu(x_2)] [1 + a_\mu(x_3)dl^\mu(x_3)] \dots \quad (5.11)$$

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

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

$$a_\mu \rightarrow U^{-1} a_\mu U + U^{-1} \partial_\mu U \quad (5.12)$$

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 analogue¹⁴ 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

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

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

¹¹For general Lie algebras, we want to write $a_\mu = a_\mu^a T_a$ where T_a are the antihermitian 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 $\text{Tr}[T_a T_b] \equiv -\frac{1}{2} \delta_{ab}$. 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.

¹²These 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.

¹³The 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.

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

¹⁵In the case of the gauge group being $SU(2)$, as mentioned in section 28.2.3, 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 S^3 (space) into 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)

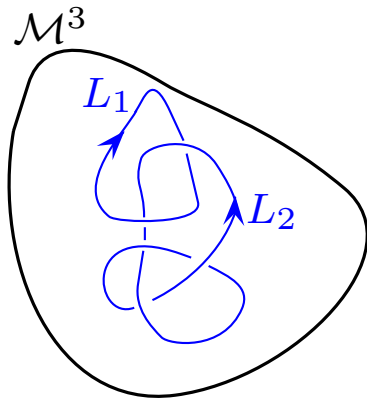


Fig. 5.1 A cartoon of a 3 manifold with a link made of two strands embedded in it.

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 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)

$$S_{CS} \rightarrow S_{CS} + 2\pi\nu k \quad (5.14)$$

where

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

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 group¹⁵.

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

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

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 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 time¹⁶. 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 28.1.1 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).

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

$$W_C(x, y) \rightarrow U(x)^{-1} W_C(x, y) U(y) \quad (5.16)$$

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. 5.16 holds. We consider

$$W_C(x, x + dx) = 1 + a_\mu dx^\mu \quad (5.17)$$

and its transformation should be

$$W_C(x, x + dx) \rightarrow U(x)^{-1} W_C(x, x + dx) U(x + dx)$$

¹⁶The 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 28.4.) In field theory language this enters the calculation by how a point-splitting regularization is implemented.

$$\begin{aligned}
&= U(x)^{-1}[1 + a_\mu dx^\mu]U(x + dx) \\
&= U(x)^{-1}[1 + a_\mu dx^\mu][U(x) + dx^\mu \partial_\mu U(x)] \\
&= 1 + [U^{-1}a_\mu U + U^{-1}\partial_\mu U]dx^\mu \quad (5.18)
\end{aligned}$$

By comparing Eq. 5.17 and Eq. 5.18 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 you don't see the metric $g_{\mu\nu}$ written anywhere in Eq. 5.13, therefore it must be metric independent. But that kind of misses the point!

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

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

which then makes it “obvious” that this is metric independent 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 reparameterization 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

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

But notice that the prefactor, including the ϵ , 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 $U(x) : \mathcal{M} \rightarrow G$ 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 ??). 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). 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!)

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 “anomalous”. 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

¹⁷This 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 ***.

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 boundary¹⁷. So to make Chern-Simons theory well-defined, you must either extend into 4d, or you can “cancel” the anomaly 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!

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

5.3.6 Chern Simons Canonical Quantization for the Abelian Case

¹⁸Note that for nonabelian Chern-Simons theories working in the $a_0 = 0$ gauge makes the a^3 term of the action vanish!

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 versa¹⁸. 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 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 Anomalies*, 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. Jolicœur 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 28.5.

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 is 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 $U(x) \in SU(2) \rightarrow S^3$. 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 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

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 accelerator¹ 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.

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

6.0.2 Which Approach?

There are several approaches to quantum gravity. While I will not make any statement about which approaches is promising, and which approaches are crazy and overpublicized², 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.

²For 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.

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 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

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

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 action³ for gravity takes the form

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

where the integration is over the entire space-time manifold \mathcal{M} , where here g is the space-time metric and R is the Ricci scalar. One might imagine that we could construct a theory of quantum gravity by plugging the Einstein-Hilbert action into the path integral form of Eq. 6.1. We obtain

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

Even without matter in the universe, the model is very nontrivial because the space-time metric can fluctuate — these fluctuations are just gravity waves⁴. 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 on 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 everywhere curvature free, still has the possibility of having a *nontrivial topology*. Thus what we are interested in is actually the different topologies that our space-time manifold might have!

³Written down first by Hilbert in 1915.

⁴Observation 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.

We thus rewrite Eq. 6.1 as

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

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

Why would we be interested in such a quantity? In short, suppose we know what the topology is of our (d -dimensional universe) at a fixed time t . We want to know the amplitudes that the topology changes as t develops. I.e., is the space-time manifold of our universe of the form $\mathcal{M} = \Sigma \times \text{time}$ 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!⁵ This connection is very roughly sketched in the chapter appendix section 6.3.

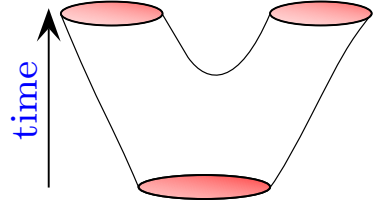


Fig. 6.1 A manifold where the topology of a space-like slice (slice at fixed time) changes as time progresses.

⁵ 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 discretizations⁶. With certain assumptions these approaches appear to be very closely related! In section 16.3 we will return to the issue of discretizing space-time and how this can result in topological gravity.

⁶ Indeed at length scales as small as the Planck length $l_{\text{Planck}} = \sqrt{\hbar G/c^3} = \hbar c/E_{\text{Planck}} \approx 1.6 \times 10^{-35} \text{m}$, there is no reason to believe space-time resembles our macroscopic idea of a smooth manifold. Roughly the ratio of the radius of the sun to the radius of an atom is 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 $\eta_{\mu,\nu} = \text{diag}[-1, 1, 1, \dots]$ 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

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

2005 and

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

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

$$\epsilon_{\mu\nu} k^\nu = 0. \quad (6.3)$$

2005

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 $x \rightarrow x - \xi$, we have

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

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

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

one obtains the Lagrangian (using differential form notation)

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

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

$$g_{\mu\nu} = e_\mu^a e_\nu^a \eta_{ab}$$

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 Lorentzian 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 GR 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 in arXiv:0706.3359, 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].
- A 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.

Topological Quantum Field Theory

7

We already have a rough picture of a Topological Quantum Field Theory (TQFT) as a quantum theory that depends on topological properties 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-manifold¹ 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-manifold². 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

¹Particularly 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 even for the analysis of even fairly simple physical systems.

²We may also allow world lines of anyons to fuse into other species as discussed in section 4.2.

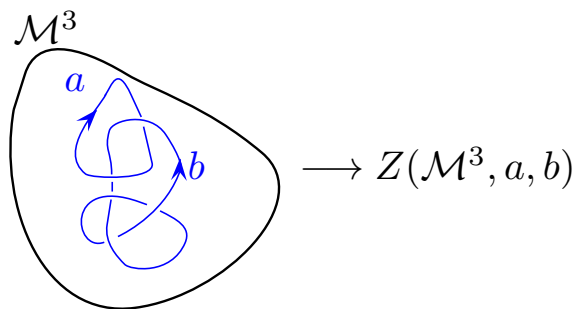


Fig. 7.1 A (2+1) dimensional TQFT takes an input of a labeled link in a manifold and produces an output of a complex number in a manner which is topologically invariant.

³Sir Michael Atiyah, a Fields medalist, was one of the foremost mathematicians of the 20th century. He specialized in geometry and topology — particularly at the interface between mathematics and physics.

⁴While 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 28.1.

⁵The 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!

⁶This 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.

is that strictly speaking, the functional integrals of Chern-Simons theory 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]³ 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 Atiyah’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 manifold⁴ 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 topology⁵ 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 space⁶. 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:

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

The reason this must be true is because $\emptyset \cup \Sigma = \Sigma$ and $\mathbb{C} \otimes V(\Sigma) = V(\Sigma)$ so the result follows⁷.

AXIOM 3: If \mathcal{M} is a $(d+1)$ -dimensional manifold with D -dimensional boundary⁸ $\Sigma = \partial\mathcal{M}$, 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 $\partial\mathcal{M}$ 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).

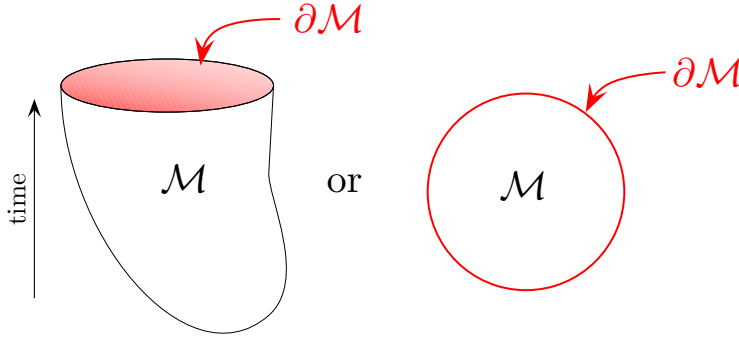


Fig. 7.2 Two depictions of a space-time manifold \mathcal{M} with boundary $\partial\mathcal{M}$. The left depiction is problematic because the only boundary of the manifold is supposed to be the red top surface $\partial\mathcal{M}$ (the black outline of \mathcal{M} really should not be there, but we 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 $\partial\mathcal{M} = \emptyset$, 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.

⁷If 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 .

⁸We use the ∂ to denote boundary. See section 28.1.4.

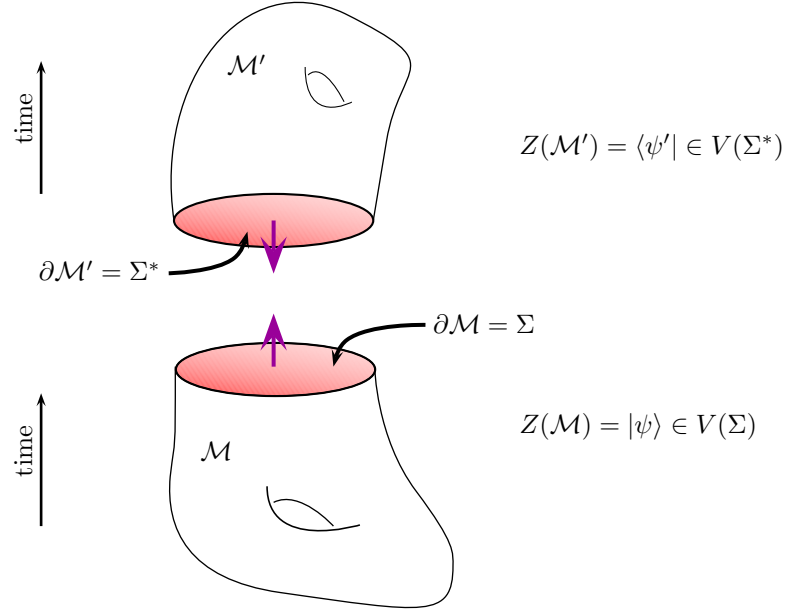


Fig. 7.3 In this picture \mathcal{M} and \mathcal{M}' are meant to fit together since they have a common boundary but with opposite orientation $\Sigma = \partial\mathcal{M} = \partial\mathcal{M}'^*$. Here $\langle\psi'| = Z(\mathcal{M}') \in V(\Sigma^*)$ lives in the dual space of $|\psi\rangle = Z(\mathcal{M}) \in V(\Sigma)$. 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 orientation 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 give⁹

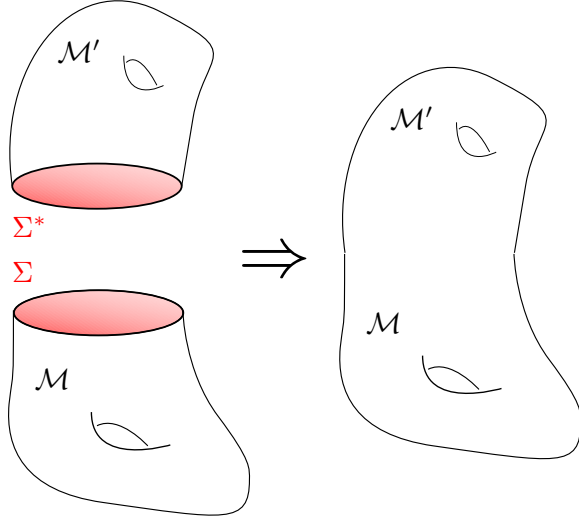
$$Z(\mathcal{M} \cup_{\Sigma} \mathcal{M}') = \langle Z(\mathcal{M}') | Z(\mathcal{M}) \rangle \quad (7.1)$$

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

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

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

⁹The notation $\mathcal{M} \cup_{\Sigma} \mathcal{M}'$ means the union of \mathcal{M} and \mathcal{M}' glued together along the common boundary Σ .



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

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

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 evolution¹⁰ similar to that shown in Fig. 7.5.

¹⁰This evolution may or may not be unitary — indeed, the dimensions of $V(\Sigma_1)$ and $V(\Sigma_2)$ may not even match if $\Sigma_1 \neq \Sigma_2$.

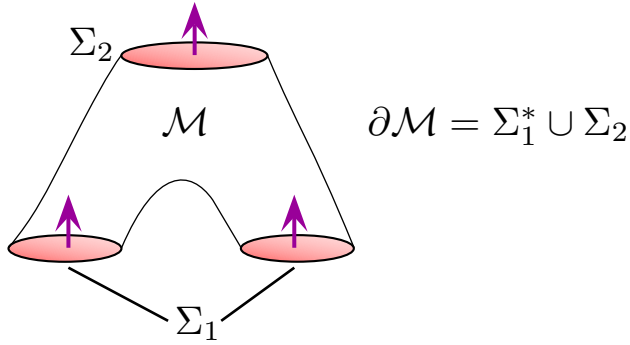


Fig. 7.5 \mathcal{M} is the cobordism between Σ_1^* and Σ_2 . I.e., $\partial\mathcal{M} = \Sigma_1^* \cup \Sigma_2$. Note that we have reversed orientation of Σ_1 here.

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

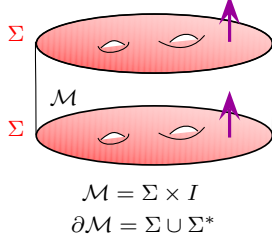


Fig. 7.6 A cobordism that can be topologically contracted to nothing acts as the identity on the Hilbert space $V(\Sigma)$.

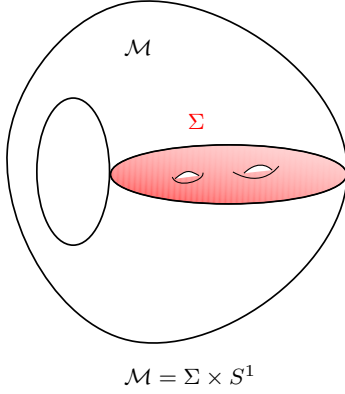


Fig. 7.7 Gluing the top of $\Sigma \times I$ to the bottom we obtain $\mathcal{M} = \Sigma \times S^1$. An important fact is that $Z(\Sigma \times S^1)$ is just the ground state degeneracy of the 2-manifold Σ .

¹¹For dimension $D > 2+1$ dimensional TQFTs we could have world-sheets of moving strings and other higher dimensional objects as well.

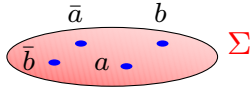


Fig. 7.8 A 2-manifold with particles in it, which are marked and labeled points. We now call the combination (the manifold and the marked points) Σ for brevity.

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

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

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

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

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

$$\text{Dim } V(S^2) = 1 \quad (7.4)$$

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.¹¹

Let us imagine that there are different anyon types which we can label as a, b, c , and so forth. The corresponding antianions 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 marked points around within the manifold (i.e., if we braid the particles around each other).

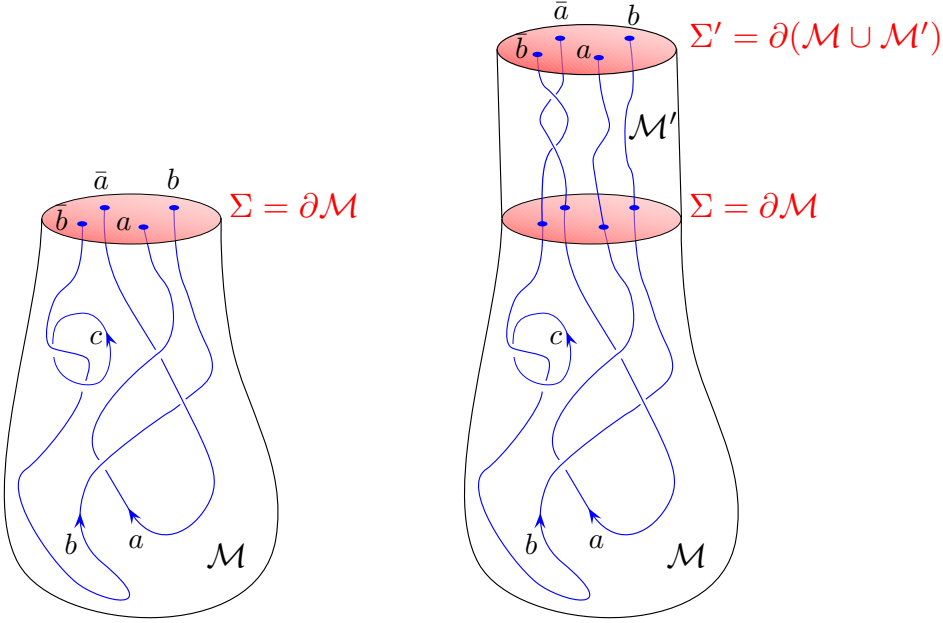


Fig. 7.9 Left: 3-manifold \mathcal{M} with particles in it, which are marked and labeled lines (the lines should be directed unless the particle is its own antiparticle). These world lines may end on the boundary $\Sigma = \partial\mathcal{M}$. The wavefunction on the boundary $\partial\mathcal{M}$ is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

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 $V(\partial\mathcal{M})$ is given by wavefunctions $|\psi_\alpha\rangle$ 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

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

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 particles¹². 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 knot¹³. The statement that it forms a nontrivial knot is a statement about the embedding of the S^1 loop in the manifold.

¹²Up 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.

¹³ D^2 is the usual notation for a two dimensional disk and S^1 again is the circle.

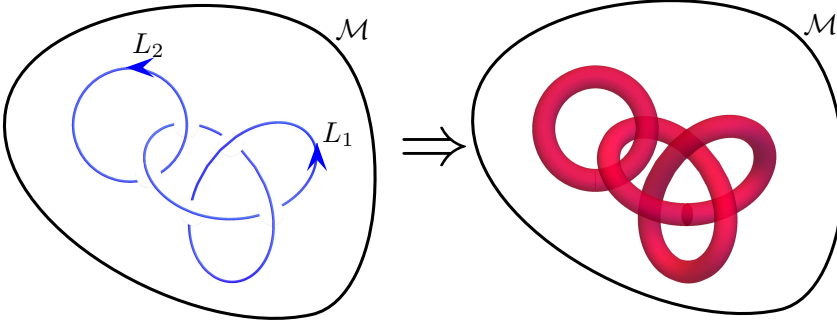


Fig. 7.10 Removing the world-lines on the left along with a thickened tube. Imagine a worm burrowing along the path of the world lines and leaving a hollow hole (colored red).

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).

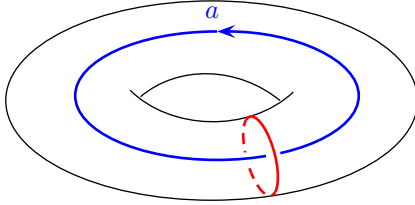


Fig. 7.11 The possible wavefunctions $|\psi_a\rangle$ that we can have on the surface of the torus can be realized by having a world-line of a particle of type a going around the handle of the torus. We can call these $Z(\text{solid torus with } a \text{ running around handle}) = |\psi_a\rangle$

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

¹⁴If you are rusty on these elementary topology manipulations, see the review in section 28.1

¹⁵Topologically 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 $B^4 = I \times I \times I \times I = D^2 \times D^2$.

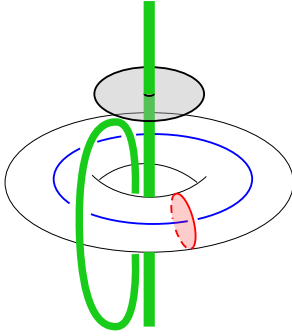


Fig. 7.12 Assembling two solid tori to make S^3 . The obviously drawn torus $D^2 \times S^1$ can be thought of as the red disk D^2 crossed with the blue circle S^1 . The remainder of space outside of this torus, including the point at infinity is the other solid torus $S^1 \times D^2$. For this “outside” 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.

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 28.1). 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 tori¹⁴

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

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 have¹⁵

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

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 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:

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

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

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

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 18 below.¹⁶

¹⁶Some 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 .

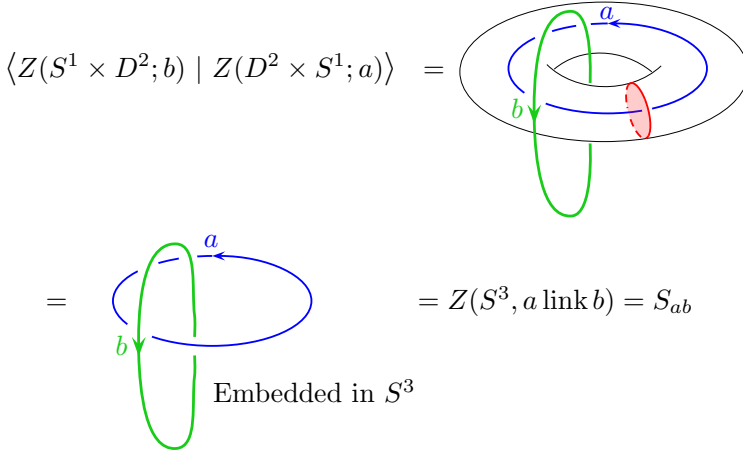


Fig. 7.13 Here we assemble a partition function for S^3 with world lines of a linking b embedded in the S^3 . To do this we glue together two solid tori each with a world line running around the handle. The green line marked b runs around the handle of the “outside” torus. The end result is known as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

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

and in fact, should be strictly less than one unless there are no nontrivial particle types 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 8-18 we will construct diagrammatic 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 $S^2 \times S^1$

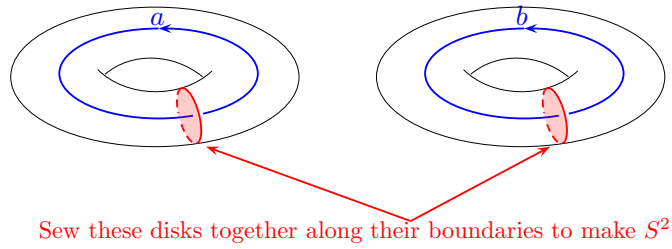


Fig. 7.14 Assembling two solid tori to make $S^2 \times S^1$. Here the two contractable disks D^2 are sewn together along their boundaries to make S^2 .

¹⁷In 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.

¹⁸One should be warned that $S^2 \times S^1$ cannot be embedded in usual three dimensional space, so visualizing it is very hard!

There is another way we can put two solid tori together to make a closed manifold¹⁷. 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 that¹⁸

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

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 together 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 have¹⁹

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

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.

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

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

¹⁹It 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.

So we have the result

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

Note that this agrees with two of our prior statements. On the one hand Eq. 7.5 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”.

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.

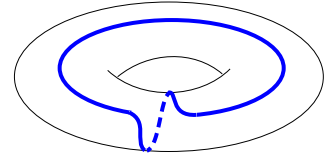


Fig. 7.15 A line that wraps both the longitude and meridian of the torus.

Chapter Summary

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

Further Reading

For discussion on the Atiyah Axioms

- M. F. Atiyah, *Proceedings of 5th Gokova Geometry and Topology Conference*, Tr. J. Mathematics, 21, 1, (1997).
<http://www.maths.ed.ac.uk/aar/papers/atiyahinttqft.pdf>
- M. F. Atiyah, *Topological quantum field theory*. Publications Mathematiques de l'IHS, 68 (1988), p. 175-186
http://www.numdam.org/item?id=PMIHES_1988_68_175_0

Part II

Hilbert Space and Planar Diagrammatic Algebra

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.

8.1 Basics of Particles and Fusion — The Abelian Case

Particle types:

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

Fusion

World lines can merge which we call fusion, or do the reverse, which we call splitting. If an a particle merges with b to give c , we write $a \times b = b \times a = c$. This is shown diagrammatically in Fig. 8.1.

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.

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 $a = 0 \dots m - 1$, 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 this¹ as 1 or 0 or I or e . The identity fuses trivially

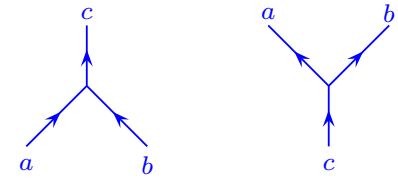


Fig. 8.1 Fusion and splitting diagrams can be thought of as part of a space-time history of the particles. If we are describing two separated particles a and b whose overall quantum number is c , we would describe the ket for this state using the right hand picture — which we can think of as a space-time description of how the current situation (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 section 10.1.

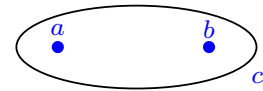


Fig. 8.2 Another notation to describe the fusion of two particle types to make a third $a \times b = c$. The two particles need not be close to each other. This figure is equivalent to the right of Fig. 8.1.

¹It 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 group of particles is fusing by addition (as we discussed in the charge-flux model) the identity should be 0. But if our group fuses by multiplication, identity is more naturally 1. See note 5 in chapter 28

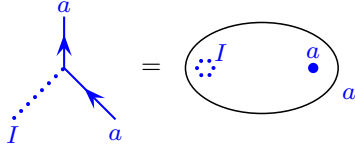


Fig. 8.3 Two depictions of fusion of a particle with the identity $a \times I = a$. On the left, the dotted line indicates the identity. On the right, the dotted circle is supposed to indicate the identity. The circle surrounding both a and I , has overall particle type a .

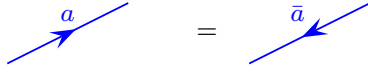


Fig. 8.4 A particle going forward should be equivalent to an antiparticle going backwards.

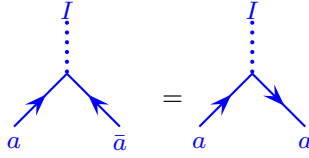


Fig. 8.5 Fusion of an anyon with its anti-anyon to form the identity can be thought of as a particle turning around in space-time.

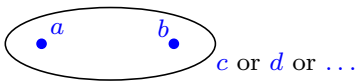


Fig. 8.6 Multiple possible fusion channels. Here we show that a and b can fuse together to give either c or d or other possible results.

$$a \times I = a$$

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.

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

$$a \times b = c + d + \dots \quad (8.1)$$

meaning that a and b can come together to form either c or d or \dots , 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 any physics (if not of physics altogether).

The quantum number (or “charge”) of a particle is locally conserved in space. Consider, for example, Fig. 8.7. On the left, a particle a propagates along and suddenly something complicated happens locally. If only a single particle comes out of this region it must also be a particle of type a . (If two particles come out of this region, we could have a split into two other species as in the right of Fig. 8.1). We sometimes call this the **no transmutation** principle. It allows us to conclude that the complicated picture on the left of Fig. 8.7 must be equal to some constant times the simple propagation of an a particle as shown on the right.

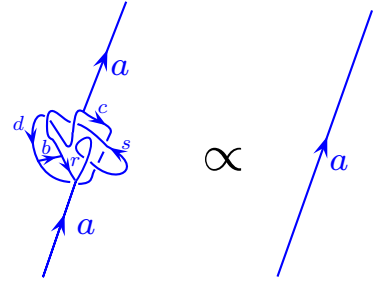


Fig. 8.7 If a particle a goes into a spacetime region, then a net particle charge a must come out. This is also sometimes called the “no-transmutation” principle. From far away, one can ignore any local processes (up to an overall constant).

If two particles (maybe far away from each other) fuse together to some 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 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 fusion channel — measuring each anyon individually does not tell you the fusion of the two!

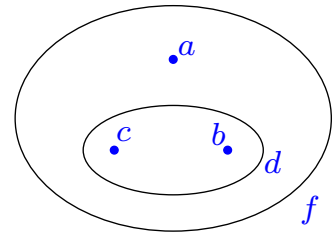


Fig. 8.8 In this picture b and c fuse to d . Then this d fuses with a to give an overall fusion channel of f . No amount of braiding b around c will change the fact that the two of them fuse to d . However, if we braid a with b and c , this can change the fusion of b with c subject to the constraint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

Antiparticles in the Case of Multiple Fusion Channels

When we have multiple fusion channels (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.

²Fibonacci, 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, ... is named after him, although it was known in India hundreds of years earlier!

³Fibonacci 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.

8.2.1 Example: Fibonacci Anyons

Perhaps the simplest nonabelian example is the anyon system known as Fibonacci² 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 27. Fibonacci anyons are closely related to the $SU(2)_3$ Chern-Simons theory³.

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

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

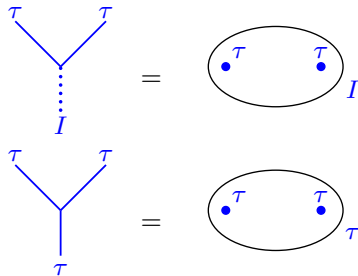


Fig. 8.9 Two different notations for the two different fusion channels of two Fibonacci anyons

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

The first two of these rules hardly need to be written down (they are implied by the required properties of the identity). It is the final rule that is 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 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 as⁴ $|N\rangle$. 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. However, as we will see below in section 17.2, 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 $|1\rangle$.

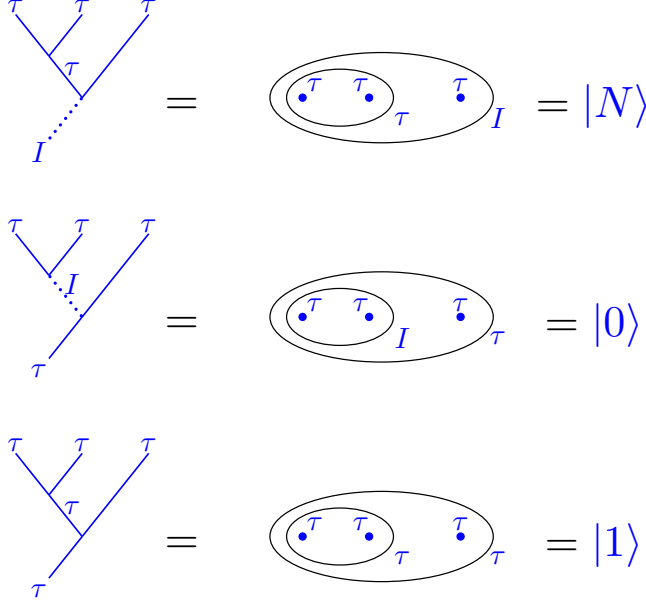


Fig. 8.10 Notations for the three different orthogonal fusion channels of three Fibonacci anyons. The notation $|N\rangle$, $|1\rangle$ and $|0\rangle$ are common notations for those interested in topological quantum computing with Fibonacci anyons!

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

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

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

8.2.2 Example: Ising Anyons

The Ising⁵ anyon system is extremely closely related to $SU(2)_2$ Chern-Simons theory⁶, and this general class of anyon is believed to be realized

⁵The 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.

⁶The 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 ***.

⁷Another 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.

in the $\nu = 5/2$ quantum Hall state (see section 27), topological superconductors, and other so-called Majorana systems (see section ***).

The Ising theory has three particle types⁷:

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

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*. 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 $d = \sqrt{2}$ (See Eq. 3.8).

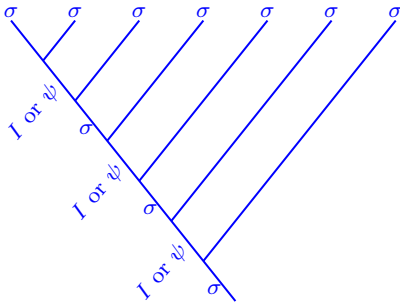


Fig. 8.11 The fusion tree for many σ particles in the Ising anyon theory.

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

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

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 $a \times b = \dots + c + \dots$, and c only occurs once on the right hand side. If c occurs more than once on the right hand side, then N_{ab}^c simply counts the number of times it occurs.

What does it mean that a particle type can occur more than once in the list of fusion outcomes? It simply means that the fusion result can occur in multiple orthogonal ways⁸ in which case a diagram with a vertex showing a and b fusing to c should also contain an index ($\mu \in 1 \dots N_{ab}^c$) 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 9.1.2. See also section ***). It is good to keep in mind that such more complicated cases exist.

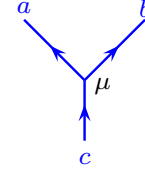


Fig. 8.12 Multiple fusion channels. In nonabelian theory fusion of a and b to c can occur in multiple orthogonal ways when $N_{ab}^c > 1$. To specify which way they fuse, we add an additional index $\mu \in 1 \dots N_{ab}^c$ at the vertex as shown.

Elementary properties of the fusion multiplicity matrices

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

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

- Time reversal

$$N_{ab}^c = N_{\bar{a}\bar{b}}^{\bar{c}} \quad (8.3)$$

- Trivial fusion with the identity

$$N_{aI}^b = \delta_{ab} \quad (8.4)$$

- Uniqueness of inverse

$$N_{ab}^I = \delta_{b\bar{a}} \quad (8.5)$$

It is sometimes convenient to define

$$N_{ab\bar{c}} = N_{ab}^c \quad (8.6)$$

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. 8.4 and Eq. 8.5 are actually the same. Further, using Eq. 8.6 along with the symmetry of N_{abc} we can derive identities such as

$$N_{ab}^c = N_{a\bar{c}}^{\bar{b}} = N_{\bar{a}b}^c. \quad (8.7)$$

where in the last step we used Eq. 8.3.

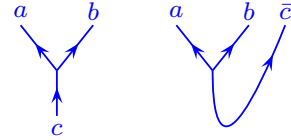


Fig. 8.13 An equivalence of N_{ab}^c with $N_{ab\bar{c}}$. Both types of vertices have the equivalent fusion multiplicity. Note that the left half of the right picture is exactly equivalent to the left — c is entering the vertex from below (then this c turns over to become a \bar{c} going up on the far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

⁸While 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

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

and the 8 occurs twice on the right.

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

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

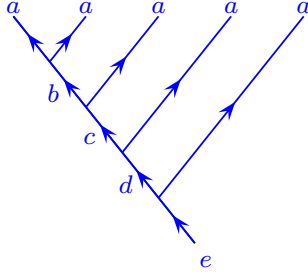


Fig. 8.14 Fusing five a type anyons together into a final result e .

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

$$d_a = \text{largest eigenvalue of } [N_a] \quad (8.8)$$

Note that this implies $d_a = d_{\bar{a}}$ given the symmetries of N .

Example of Fibonacci Anyons

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

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

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 $d_\tau = (1 + \sqrt{5})/2$, in agreement with Eq. 8.2.

Example of Ising Anyons

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

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

where the first row and column represent the identity, the second row and column represent σ and the third row and column represent ψ . So, for example, the second row here indicates that $\sigma \times \sigma = I + \psi$. Again, it is an easy exercise to check that the largest eigenvalue of this matrix is $d_\sigma = \sqrt{2}$ as described in section 8.2.2.

8.3.1 Associativity

It should be noted that the fusion multiplicity matrices N are 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

$$\sum_d N_{ab}^d N_{cd}^e = \sum_f N_{cb}^f N_{af}^e \quad (8.9)$$

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

$$[N_a, N_c] = 0 \quad (8.10)$$

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. 8.7). A set of normal matrices that all commute can be simultaneously diagonalized, thus

$$[U^\dagger N_a U]_{xy} = \delta_{xy} \lambda_x^{(a)} \quad (8.11)$$

and all N_a 's get diagonalized with the same unitary matrix U . Surprisingly (as we will see below in section 18) for well behaved (so-called “modular”⁹ anyon theories) the matrix U is precisely the modular S -matrix we discussed above in Eq. 7.6 !

8.4 Fusion and Hilbert Space

The structure of fusion rules can be used to calculate the ground state degeneracy of wavefunctions on 2-dimensional manifolds¹⁰. 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,

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

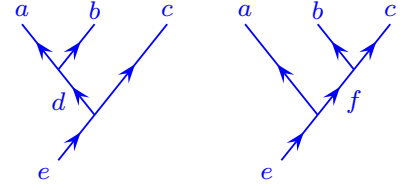


Fig. 8.15 Fusing $(a \times b) \times c$ should be equivalent to $a \times (b \times c)$. On the left a and b fuse to d first then this composite fuses with c to give e . On the right b and c fuse to f first, then this composite fuses with a to give e . Both diagrams represent the same physical Hilbert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

⁹For nonmodular theories, we can still diagonalize N in the form of Eq. 8.11, and the resulting unitary matrix U is sometimes known as the *mock S*-matrix.

¹⁰We are again assuming manifolds are always orientable – so this excludes objects like the Klein bottle or the Möbius strip.

¹¹By nontrivial we mean this particle is not the vacuum particle.

¹²For 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. 8.13.

¹³It 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.

¹⁴Since 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.

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 (nontrivial¹¹) 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

$$\text{Dim } V(S^2 \text{ with one (nontrivial) anyon}) = 0 \quad (8.12)$$

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.¹²

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 have¹³

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

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 type¹⁴. 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

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

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 19). 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

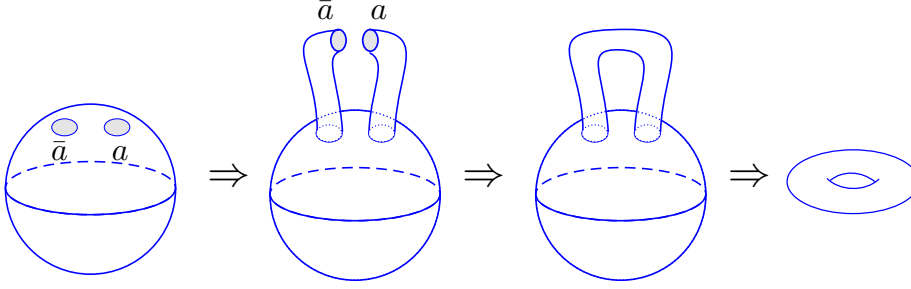


Fig. 8.16 Surgering the twice punctured sphere into a torus. This is the gluing axiom in action. Note that we are implicitly assuming the system is trivial in the “time” direction, which we assume to form a circle S^1_{time} .

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 other¹⁵.

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.

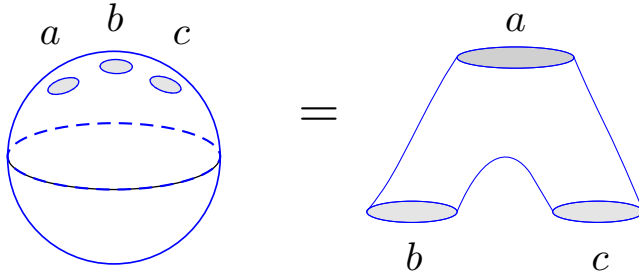


Fig. 8.18 A three-times punctured sphere is known as a “pair of pants”.

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 $a \times b \times c \rightarrow I$ (for the bottom pair of pants) and $\bar{a} \times \bar{b} \times \bar{c} \rightarrow I$ (for the top pair of pants). This number of

¹⁵In 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 $\text{Tr}[Z((S^1 \times I) \times S^1_{time})] = Z(T^2 \times S^1_{time})$.

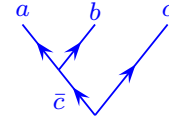


Fig. 8.17 Three particles that fuse to the identity. There are $N_{abc} = N_{ab}^{\bar{c}}$ different fusion channels.

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

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

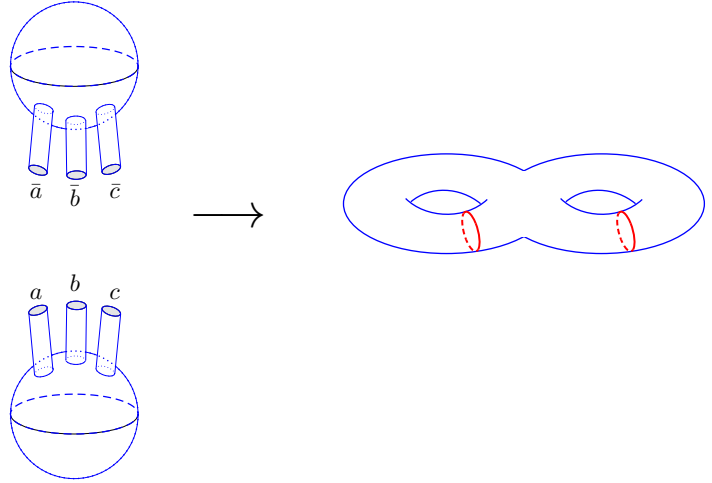


Fig. 8.19 Sewing together two pair of pants to form a two-handled torus.

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. 8.12) one can have a single anyon on a torus (See note 12 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

$$\text{Dim } V(T^2 \text{ with one } a) = \sum_b N_{b\bar{b}a} \equiv L_a \quad (8.13)$$

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

$$\text{Dim } V(\text{Three handled Torus}) = \sum_{abc} L_a L_b L_c N_{\bar{a}\bar{b}\bar{c}} \quad (8.14)$$

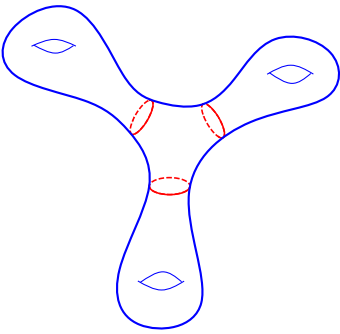


Fig. 8.20 Decomposing a three handled torus into three copies of a torus with puncture (the puncture is the red collar), and a single pants in the middle. I have resisted the urge to draw a three handled object as being covered with moss.

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. 8.14 to obtain

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

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

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

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

Exercise 8.2 Fusion and Ground State Degeneracy

To determine the ground state degeneracy 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 $\mathcal{D}^2 = \sum_a d_a^2$.

Change of Basis and F -symbols¹

As mentioned in Fig. 8.15, one can describe the space of three particles in two different ways. If we are considering the space spanned by the fusion of $a \times b \times c$ as in the figure, we can describe the space by how a fuses with b (the value of d on the left of the figure), or by how b fuses with c (the value of f in the figure). Either of these two descriptions should be able to describe the space, but in different bases. We define the change of basis as a set of unitary matrices^{2,3,4} called F , as shown in Fig. 9.1.

$$\text{Diagram 1} = \sum_f [F_e^{abc}]_{df} \text{Diagram 2}$$

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 e . For fixed a, b, c and e , the matrix F is unitary in its subscripts d, f . One often uses the convention that the F matrix is defined to be zero if the fusion diagram is not allowed, i.e. if any of the fusion multiplicities N_{ab}^d , N_{dc}^e , N_{bc}^f , N_{af}^e are zero.

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.

9.0.1 Example: Fibonacci Anyons

Again we turn to the example of Fibonacci anyons for clarification. We imagine fusing together three τ particles. As shown in Fig. 8.10, there is a single state $|N\rangle$ in which the three fuse to the identity I . It should not matter if we choose to fuse the leftmost two anyons first, or the rightmost two. In either case there is only one possible state for the outcome. We can thus draw the simple identity shown in Fig. 9.2. 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

¹This 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 appendices.

²In cases where there are fusion multiplicities N_{ab}^c greater than one (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 9.1.2

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

⁴Typically if a or b or c is the identity, then we choose a gauge so that F is the identity as well. See appendix 9.1.3.

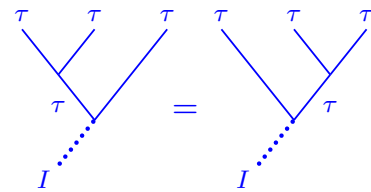


Fig. 9.2 There is only one state in the Hilbert space of three Fibonacci anyons fusing to the identity. Thus it does not matter if you fuse the left two first or the right two first, you are describing the same state.

$$\begin{array}{c} \tau \\ \swarrow \quad \searrow \\ \tau \quad \tau \end{array} \begin{array}{c} \tau \\ \swarrow \quad \searrow \\ \tau \quad \tau \end{array} = \begin{array}{c} \tau \quad \tau \quad \tau \\ \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \end{array} = |0\rangle$$

$$\begin{array}{c} \tau \\ \swarrow \quad \searrow \\ \tau \quad \tau \end{array} \begin{array}{c} \tau \\ \swarrow \quad \searrow \\ \tau \quad \tau \end{array} = \begin{array}{c} \tau \quad \tau \quad \tau \\ \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \end{array} = |1\rangle$$

Fusing the two particles on the left first

$$\begin{array}{c} \tau \\ \swarrow \quad \searrow \\ \tau \quad \tau \end{array} \begin{array}{c} \tau \\ \swarrow \quad \searrow \\ \tau \quad \tau \end{array} = \begin{array}{c} \tau \quad \tau \quad \tau \\ \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \end{array} = |0'\rangle$$

$$\begin{array}{c} \tau \\ \swarrow \quad \searrow \\ \tau \quad \tau \end{array} \begin{array}{c} \tau \\ \swarrow \quad \searrow \\ \tau \quad \tau \end{array} = \begin{array}{c} \tau \quad \tau \quad \tau \\ \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \end{array} = |1'\rangle$$

Fusing the two particles on the right first

Fig. 9.3 Two ways to describe the same two dimensional space. The basis $\{|0\rangle, |1\rangle\}$ fuses the left two particles first, whereas the basis $\{|0'\rangle, |1'\rangle\}$ fuses the right two particles first. Again note that we are considering kets so the tree branches point upwards.

the 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. 9.3. 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. 9.3.

The space of states spanned by the three anyons is the same in either description. Thus, there must be a unitary basis transform given by

$$\begin{pmatrix} |0\rangle \\ |1\rangle \end{pmatrix} = \begin{pmatrix} F_{00'} & F_{01'} \\ F_{10'} & F_{11'} \end{pmatrix} \begin{pmatrix} |0'\rangle \\ |1'\rangle \end{pmatrix} \quad (9.1)$$

⁵We can redefine kets with different gauge choices (see section 9.1.3) and this will insert some phases into the off-diagonal of this matrix, but the simplest gauge choice gives the matrix as shown.

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 by⁵

$$F = \begin{pmatrix} \phi^{-1} & \phi^{-1/2} \\ \phi^{-1/2} & -\phi^{-1} \end{pmatrix} \quad (9.2)$$

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 the next section (Section 9.0.2) we will discuss how this matrix is derived.

It is important to emphasize that the F -matrix is the same even if one of the anyons charges being fused is actually a cluster of several anyons. For example, in Fig. 9.4, this is precisely the same transformation as in Eq. 9.1, but we must view the cluster of two anyons on the left (red), which fuse to τ as being a single τ particle.

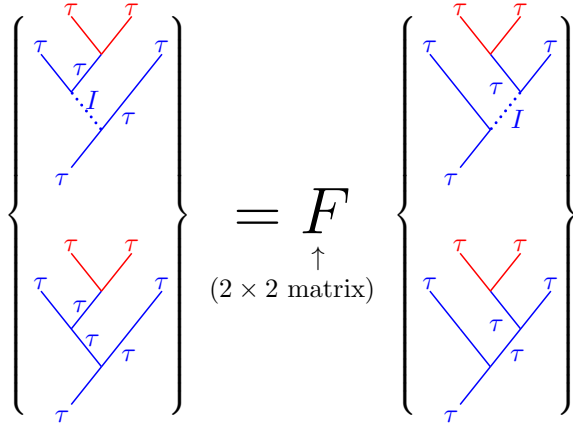


Fig. 9.4 The F -matrices are the same even if one of the anyon charges is made up of a cluster of other anyons. In this particular picture, the cluster of two anyons on the very top (in red) has overall charge τ . If we replace this cluster of two anyons with just a single τ , this would be precisely the same transformation as in Eq. 9.1.

9.0.2 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. I.e., 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. 9.4. 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. 9.5 (the horizontal step on the bottom is equivalent to that shown in Fig. 9.4). Indeed, in this figure one sees that one can go from the far left to the far right of the diagram via two completely different paths (the top and the bottom path) and the end result on the far right should be the same either way. This diagram, known as the pentagon diagram⁶, puts a very strong constraint on the F matrices, which written

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

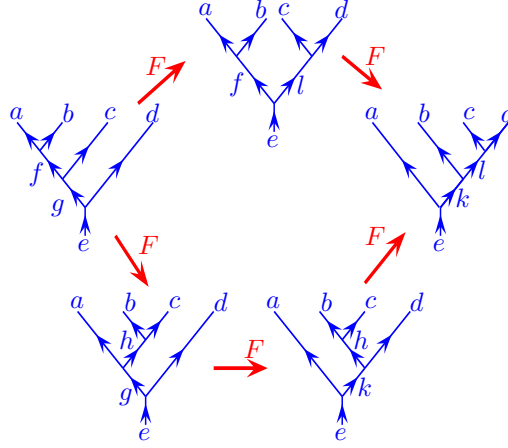


Fig. 9.5 Pentagon Diagram. Each step in the diagram is a new description of the same basis of states via and F -matrix.

out algebraically would be

$$[F_e^{fcd}]_{gl}[F_e^{abl}]_{fk} = \sum_h [F_g^{abc}]_{fh}[F_e^{ahd}]_{gk}[F_k^{bcd}]_{hl} \quad (9.3)$$

where the left hand side represents the top route of the figure and the right hand side represents the bottom route.⁷

⁷It is very worth working through this to make sure you understand how this equation matches up with the figure!

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 9.1.3). See exercise 9.1. Further, for *any* given set of fusion rules there are a finite set of possible solutions of the pentagon equation — a property that goes by the name “rigidity” Etingof et al. [2015].

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

9.1 Appendix: F -matrix Odds and Ends

9.1.1 Unitarity of F

The F -matrix relation we defined as

$$\begin{array}{c} a \quad b \quad c \\ \diagdown \quad \diagup \quad \diagup \\ d \quad e \end{array} = \sum_f [F_e^{abc}]_{df} \begin{array}{c} a \quad b \quad c \\ \diagdown \quad \diagup \quad \diagup \\ e \quad f \end{array}$$

The fact that F is unitary in its indices d and f means we can also

write

$$\text{Diagram with lines } a, b, c \text{ entering and } e \text{ exiting, labeled } f = \sum_d [F_e^{abc}]^* \text{Diagram with lines } a, b, c \text{ entering and } e \text{ exiting, labeled } d$$

9.1.2 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. 9.6.

$$\text{Diagram with lines } a, b, c \text{ entering and } e \text{ exiting, labeled } \mu, \nu = \sum_{f, \alpha, \beta} [F_e^{abc}] (d \mu \nu)(f \alpha \beta) \text{Diagram with lines } a, b, c \text{ entering and } e \text{ exiting, labeled } \alpha, \beta$$

Fig. 9.6 The F -matrix equation with fusion multiplicities greater than one. Here the vertex indices are $\mu \in 1 \dots N_{ab}^d$ and $\nu \in 1 \dots N_{dc}^e$ and $\alpha \in 1 \dots N_{bc}^f$ and $\beta \in 1 \dots N_{af}^e$.

9.1.3 Gauge Transforms and the F -matrix

We have the freedom to make gauge transformations on our diagrams. In particular the vertices can be multiplied by a phase as shown in Fig. 9.7

$$\text{Diagram with lines } a, b, c \text{ entering and } c \text{ exiting, labeled } \mu \rightarrow u_{c;\mu}^{ab} \text{Diagram with lines } a, b, c \text{ entering and } c \text{ exiting, labeled } \mu$$

Fig. 9.7 We have the freedom to make a gauge transform of a vertex by multiplying by a phase $u_{c;\mu}^{ab}$.

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

$$[F_e^{abc}]_{(d\mu\nu)(f\alpha\beta)} \rightarrow \frac{u_{e;\beta}^{af} u_{f;\alpha}^{bc}}{u_{d;\mu}^{ab} u_{e;\nu}^{dc}} [F_e^{abc}]_{(d\mu\nu)(f\alpha\beta)} \quad (9.4)$$

or more simply in cases where there are no indices at vertices ($N_{ab}^c \leq 1$), we have

$$[F_e^{abc}]_{df} \rightarrow \frac{u_e^{af} u_f^{bc}}{u_d^{ab} u_e^{dc}} [F_e^{abc}]_{df} \quad (9.5)$$

As we shall see in section 10.4, 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. 9.7) 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. Thus we should always use $u_a^{aI} = u_a^{Ia} = 1$. Nonetheless, if there are cases where we do want to specify that a vacuum line has branched off at one particular point, we then do have the option of choosing a nontrivial gauge (See Lin and Levin [2014] for further discussion of this possibility).

Chapter Summary

- This is

Further reading

This is some reading.

Exercises

⁸Strictly speaking when there are fusion multiplicities, $N_{bc}^a > 1$, then one also needs an additional index at each vertex.

Exercise 9.1 Fibonacci Pentagon

In a TQFT (indeed, in any tensor category), a change of basis is described by the F-matrix as shown⁸ in figure 9.1. Consistency of F-matrices is enforced by the pentagon equation (Fig. 9.5 and Eq. 9.3).

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 choice. Using the pentagon equation determine the F -matrix. (To get you started, note that in Fig. 9.5 the variables a, b, c, d, e, f, g, h can only take values I and τ . You only need to consider the cases where a, b, c, d are all τ).

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

Exercise 9.2 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.

Planar Diagrams¹

10

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 digrams where particles come together and fuse, or split apart) and get an output which is a complex number as desired in Fig. 7.1. Having described the idea of the F -matrix we can begin to construct these rules. In this chapter we will focus only on planar diagrams — i.e., we do not allow lines to cross over and under each other forming braids. We can roughly 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. 10.1. 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 convenient in different contexts. We will start with a more “physics” oriented normalization in section 10.2, but we switch to a more topologically oriented normalization in section 10.3 and in later chapters.

¹This chapter develops the diagrammatic algebra in some detail. For those who would like an easier (albeit not as general) introduction to diagrammatic algebra, go to chapter 11.

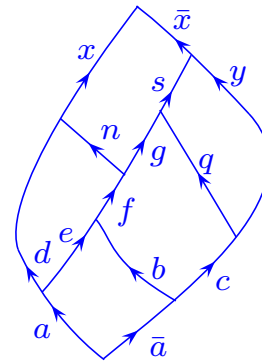


Fig. 10.1 A planar fusion diagram starting and ending at the vacuum.

10.1 What We Mean By a Diagram

If, like Fig. 10.1, 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, as we have been doing in chapters 8 and 9, 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 vacuum². We can view these diagrams with loose ends as being part 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.

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

²Many 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.

³Analogous 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. 10.1 which starts and ends in the vacuum.

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 operator³. An example of this is shown in Fig. 10.2.

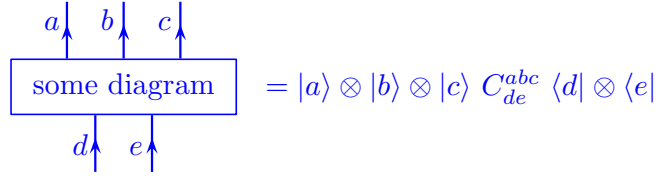


Fig. 10.2 Interpreting a part of a diagram as an operator. Incoming lines from the bottom correspond to bras and outgoing lines towards the top correspond to kets. The value of the constant C_{de}^{abc} depends on the particular diagram in the box.

An important principle is that we can construct the hermitian conjugate of a diagram with the following procedure: reflect the diagram across the horizontal axis, and also reverse all the arrows (so up-pointing arrows continue to point up). An example of this is shown in Fig. 10.3.

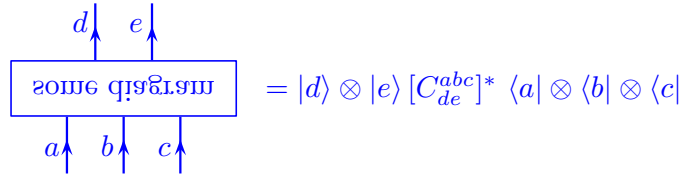


Fig. 10.3 Reflecting a diagram along a horizontal axis and then flipping the direction of all arrows gives the hermitian conjugate. This figure gives the Hermitian conjugate of the diagram shown in Fig. 10.2. We switch bras for kets and complex conjugate the constant C_{de}^{abc} .

A particularly important diagram with ends pointing both up and down is given by a simple labeled straight line, which should be interpreted as the identity operator for the particle label type as shown in Fig. 10.4.

$$a \uparrow \downarrow = |a\rangle \langle a|$$

Fig. 10.4 A labeled straight line is just an identity operator for the particle type.

10.2 Diagram Rules with Physics Normalization

To make more sense of these diagrams we should think in terms of the anyon Hilbert spaces defined in chapter 8. If we start from the vacuum at the bottom and draw a fusion tree, we construct a particular ket state

in the Hilbert space. Let us define two particular states of the Hilbert space as shown in Fig. 10.5.



Fig. 10.5 Fusion trees in the Hilbert space form an orthonormal set when we use physics normalization. The inner product of these two states is $\delta_{aa'}\delta_{bb'}\delta_{cc'}\delta_{dd'}\delta_{ee'}\delta_{ff'}\delta_{f\bar{d}}$. The final factor of $\delta_{f\bar{d}}$ is included because at the bottom of the diagram two particles fuse to the vacuum. In the isotopy invariant normalization of section 10.3 and later chapters the inner product obtains an additional factor of $\sqrt{d_a d_b d_c d_d}$.

When we draw fusion trees starting from the vacuum, and if we use the “physics” normalization of diagrams, fusion trees are defined to be orthonormal with respect to all particle labels⁴. So for example, in reference to Fig. 10.5 we have

$$\langle \text{state}' | \text{state} \rangle = \delta_{aa'}\delta_{bb'}\delta_{cc'}\delta_{dd'}\delta_{ee'}\delta_{ff'}\delta_{f\bar{d}} \quad (10.1)$$

where the final factor of $\delta_{f\bar{d}}$ is included because at the bottom of the diagram two particles fuse to the vacuum. A more convenient way to describe Eq. 10.1 and Fig. 10.5 is to take the hermitian conjugate of $|\text{state}'\rangle$ by flipping the diagram over (as in Fig. 10.3), so that we obtain Fig. 10.6.

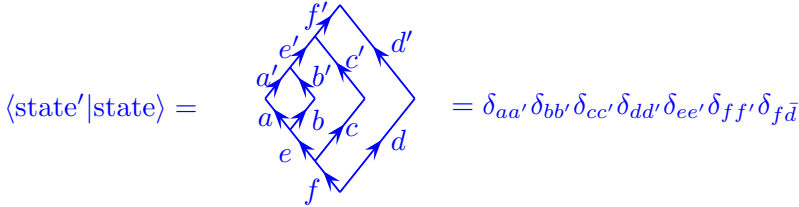


Fig. 10.6 Expressing the inner product of Eq. 10.1 as a single diagram. Again, the right hand side uses physics normalization. In isotopy invariant normalization of section 10.3 and later chapters the right hand side is multiplied by $\sqrt{d_a d_b d_c d_d}$.

Note again that the normalization of these diagrams will be changed in section 10.3 below and in later chapters.

The result on the right hand side of Fig. 10.6 is given by the principle of orthonormality of fusion trees. The result in Fig. 10.6 is also consistent with the principle of locality as shown in Fig. 10.7

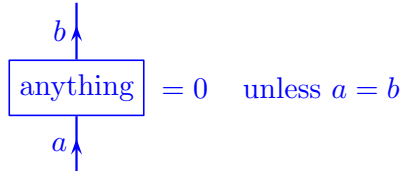


Fig. 10.7 The locality principle as in Fig. 8.7.

⁴To be precise, in cases where there are fusion multiplicities greater than one, each vertex gets an additional index (See Fig. 8.6 and sections 9.1.2 and 10.6.1). In this case, the diagrams are also orthonormal in these indices as well. For simplicity of notation we will often not write these vertex indices.

which we mentioned previously in Fig. 8.7. For example, locality implies in Fig. 10.6 that $a = a', b = b', c = c', d = d', e = e', f = f'$.

Note that the principle of orthonormality of trees implies the useful result that a loop, as shown in Fig. 10.8 is given the value of 1. At the risk of being repetitive we once again note that we will change this normalization in section 10.3 below and in later chapters, although it is correct for this section.

$$\begin{aligned} \left| \begin{array}{c} \swarrow \searrow \\ \nwarrow \nearrow \end{array} \right\rangle^a &= |\text{state}\rangle \\ \langle \text{state} | \text{state} \rangle &= \begin{array}{c} \swarrow \searrow \\ \nwarrow \nearrow \end{array}^a = 1 \end{aligned} \quad \text{Physics Normalization}$$

Fig. 10.8 The orthonormality of trees implies a particle loop gets a value of 1 if we are using physics normalization.

⁵In cases where $N_{ab}^c > 1$ we have an additional index μ at the vertex (See Fig. 8.12). In this case we must sum over the index μ as well, and it must be the same index $|c\mu\rangle\langle c\mu|$ at both vertices. See section 10.6.1.

Another very important diagram is the insertion of a complete set of states into two particle world lines as shown in Fig. 10.9. We should see this as a ket on top of a bra $\sum_c |c\rangle\langle c|$ which is our usual idea of a complete set. Another way of thinking about this is that a and b must fuse to some c , and if you sum over all possible channels c , you have summed over a complete set⁵.

$$\begin{array}{c} | \\ | \end{array} \begin{array}{c} a \\ b \end{array} = \sum_c \begin{array}{c} a \searrow \\ \swarrow b \\ c \end{array}$$

Fig. 10.9 Insertion of a complete set of states with physics normalized diagrams.

⁶Again if $N_{ab}^c > 1$ there are additional indices at the vertices and these must match as well. See section 10.6.1.

One more useful diagram is given by Fig. 10.10 which is implied by the orthogonality of tree states⁶, but it is also consistent with the locality principle shown in Fig. 10.7.

$$\begin{array}{c} \swarrow \searrow \\ \nwarrow \nearrow \end{array} \begin{array}{c} d \\ b \\ a \\ c \end{array} = \delta_{cd} \begin{array}{c} | \\ | \end{array}$$

Fig. 10.10 This identity is implied by the orthogonality of trees with the same branching structure, but it is also a result of the locality principle.

10.2.1 Causal Isotopy (vs. Full Isotopy)

Keeping with the idea of diagrams that are planar (no over- and under-crossings), we now consider how we may deform these diagrams. We need to ask how much topological invariance we should really expect from our 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.

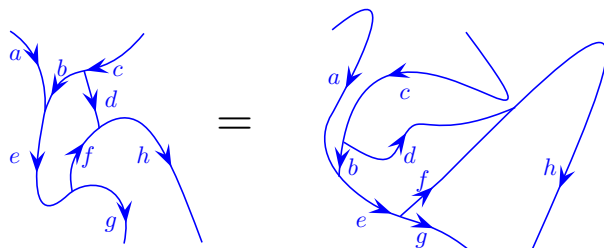


Fig. 10.11 For a theory with full two-dimensional isotopy invariance, these two diagrams should evaluate to the same result. Full isotopy invariance allows us to distort the diagram in any way as long as we do not cut any strands or cross lines through each other.

In chapter 11 we will consider a subset of theories which have a much higher level of topological invariance, known as full isotopy invariance, which allows us to freely distort diagrams in either the space or time direction and further allow us to interchange the two.⁷

In this chapter we do not allow full isotopy invariance but rather assume only what we call *causal isotopy*⁸. Here we allow deformation of space-time diagrams so long as we do not change the time-direction motion of any particles. 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. 10.12.

⁷Those who would like to minimize theoretical complexity might want to skip the rest of this chapter and go directly on to chapter 11 and follow the rules presented there. One should just realize that the theories presented there are not generic.

⁸This is not standard nomenclature.

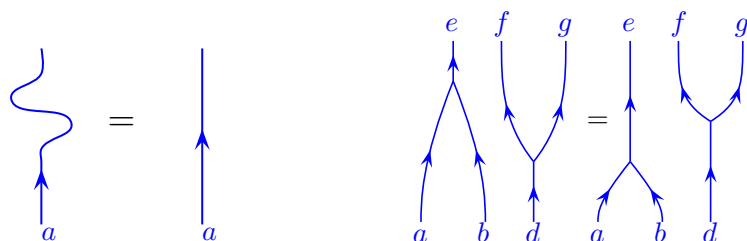


Fig. 10.12 Two examples of deformations that are allowed under causal isotopy. Deformations of the path are allowed as long as they do not require a particle to reverse directions in the time-like direction. In the left example, this deformation is allowed because in both cases the particle continues to move forward in the time direction. In the right example, the temporal order of the vertices does not matter.

Certain deformation of diagrams are not allowed by causal isotopy. Two examples of such disallowed deformations are given in Fig. 10.13. 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. 10.13 a vertex is altered so instead of a b particle coming in to the vertex, a \bar{b} particle goes out. In this case we must have a a with \bar{b} annihilation event in the right of the diagram that does not exist in the simpler diagram where a and b directly fuse to c . Thus these two diagrams do not evaluate to the same result.

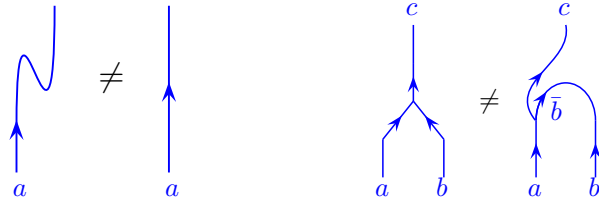


Fig. 10.13 Two examples of disallowed transformations under causal isotopy.

10.2.2 Summary of Planar Diagram Rules in Physics Normalization

⁹Any operator can be defined completely by how it acts on a basis of states. Similarly we can define the action of a diagram with lines pointing upwards and downwards, as in Fig. 10.2, by the complex amplitudes it gives when placed within some known bra on top and ket on the bottom which yield a complete diagram with no lines that end either upwards or downwards.

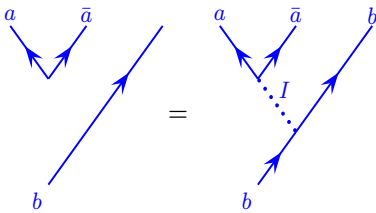


Fig. 10.14 One can always add or remove the identity (or vacuum) line to any diagram.

¹⁰This includes particle creation and annihilation as a special case where two particles fuse to the vacuum.

With the principles we have now discussed we should be able to evaluate any planar diagram — taking a space-time process and turning it into an amplitude (i.e, a complex number). Similarly we can use the same principles to evaluate operators⁹ such as Eq. 10.2.

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 10.2.1. 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. 10.14.
- (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. For example, this tells us immediately in Fig. 10.6 that $d = d'$ and $c = c'$ and $b = b'$ and $f = f'$. Only the vacuum line is allowed to abruptly terminate.
- (5) Splitting and fusion vertices are allowed¹⁰ for multiplicities $N_{ab}^c > 0$ (See section 8.3).
- (6) A hermitian conjugate is given by reflection of a diagram around a horizontal line along with flipping the direction of arrows as shown in Fig. 10.3.
- (7) One can use F -moves to change the structure of fusion trees in order to simplify their structure. For example, in Fig. 10.15, the

diagram on the left is turned into the one on the right using an F -move.

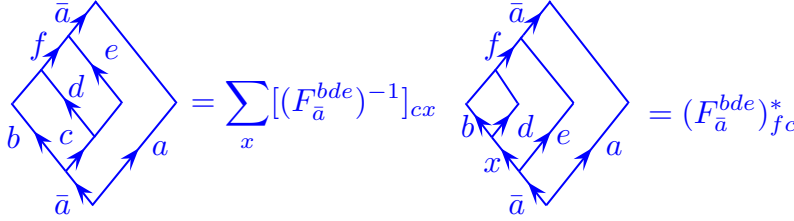


Fig. 10.15 The diagram on the left is evaluated by applying an F -move to the left half of the diagram. The resulting diagram is evaluated to delta functions as in Fig. 10.6, in particular enforcing $x = f$. Finally we use the unitarity of F in the last step. This diagram is evaluated with physics normalization.

- (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. 10.15) we can use the orthonormality of trees to complete the evaluation (as in Fig. 10.6).

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 into a complex number. The structure we have defined thus far (our Hilbert space and F matrices) is known as a unitary fusion category. Many systems may have additional structure (such as braiding rules) which we have not yet mentioned.

10.3 Seeking Full Isotopy Invariance: Isotopy Invariant Normalization

We would very much like the diagrammatic rules of our topological theories to obey full isotopy invariance — meaning that any smooth deformation of a diagram leaves its value unchanged as indicated in Fig. ???. In many cases we can make some small changes to normalizations to remove some impediments to full isotopy invariance.

Let us first examine where the most obvious problem lies. For a nice topological theory (meaning one with full isotopy invariance) we would want to have the so-called zig-zag identity shown in Fig. 10.16 (which is not a property of theories having only causal isotopy invariance as mentioned in Fig. 10.13).

Unfortunately, a set of F matrices (even if they satisfy the pentagon self-consistency condition Eq. 9.3) does not generically satisfy this zig-zag identity Fig. 10.16. To see this, consider the manipulations shown in Fig. 10.17. With the physics normalization of diagrams we have been using, the zig-zag identity does not hold.

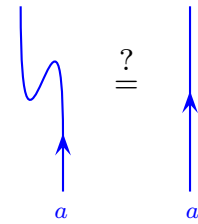


Fig. 10.16 A topological theory with full isotopy invariance should have this “zig-zag” identity. However, generically a set of F matrices will not satisfy this equality (See Fig. 10.17). We can often repair this problem by changing the normalization of kets.

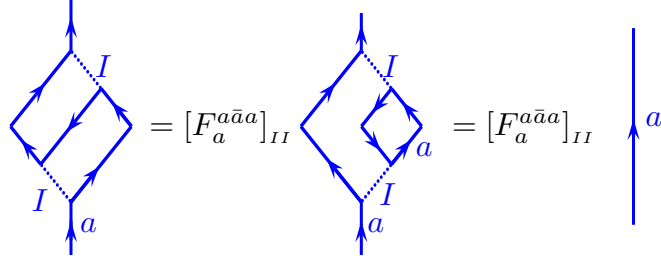


Fig. 10.17 Straightening a zig-zag wiggly line incurs a factor of F using physics normalization of diagrams. In the first step we use an F -move on the lower part of the diagram. We then use orthogonality of the tree to remove the small a bubble. This part of the diagram is just Fig. 10.8. Thus this small a bubble can be removed. We conclude that with the physics normalization we cannot satisfy the zig-zag identity Fig. 10.16.

To fix this problem, we take a cue from the Kauffman bracket invariant and change our definition of diagrams just by a small bit. In particular, let a simple loop of particle a , as shown in Fig. 10.18, be given a value of d_a . This is different from our prior definition where we set the loop value to one as in Fig. 10.8. The change here only means that we will be working with unnormalized bras and kets. We will call this normalization “isotopy normalization”.

$$\begin{aligned}
 \left| \begin{array}{c} \text{zig-zag} \\ a \end{array} \right\rangle &= |\text{state}\rangle \\
 \langle \text{state} | \text{state} \rangle &= \text{diamond with } a = \text{circle with } a = d_a \quad \text{Isotopy Normalization}
 \end{aligned}$$

Fig. 10.18 Using a new normalization (which we call “isotopy normalization”) of bras and kets. Compare to Fig. 10.8.

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 expectations 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.

With this new normalization, we can recalculate the value of a zig-zag wiggly line analogous to that of Fig. 10.16.

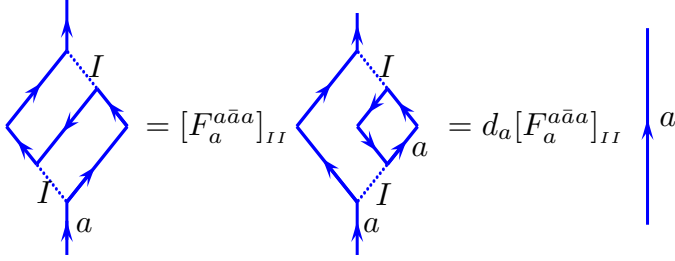


Fig. 10.19 With the new isotopy invariant normalization of diagrams, straightening a zig-zag wiggly incurs a factor of $d_a[F_a^{a\bar{a}a}]_{II}$. We will choose the value of d_a so as to make this factor into unity if we can.

Let us choose¹¹

$$d_a = \frac{1}{[F_a^{a\bar{a}a}]_{II}} \quad (10.2)$$

For the time being, let us assume that this definition gives d_a real and positive¹² and we will return to this issue in section 10.4.

Given the choice of Eq. 10.2, we see from Fig. 10.19 that the wiggly lines (like the left of Fig 10.16) can be straightended out freely. As we will see in chapter 12, the normalization constant d_a will turn out¹³ to be the same quantum dimension d_a that we found in Eq. 8.8 from the Hilbert space dimension of fusing anyons together.

In this section we are going to try to massage our theory into a form that is much closer to isotopy invariance. In some cases we will succeed in our quest of achieving isotopy invariance, and in other cases we will fall short and we will need to add further complications to our theory.

Having changed the normalization of our kets, for consistency we need to change the normalization of fusions and splittings to match. Following suit, we define new normalization of vertices as shown in Fig. 10.20.

Fig. 10.20 New Normalization for vertices. Note that this is consistent with Fig. 10.8 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. 10.6 should now have a factor of $\sqrt{d_a d_b d_c d_d}$ on the right hand side. Similarly our completeness diagram Fig. 10.9 and our bubble diagram Fig. 10.10 need to be modified as shown in Fig. 10.21 and 10.22¹⁴

¹¹Note that some references, including Bonderson [2007] and Kitaev [2006] define d_a to be the absolute value of this quantity so it is always positive.

¹²We will see that for non-self dual particles $a \neq \bar{a}$ we can always make a gauge choice so that this definition of d_a is real and positive. For self-dual particles $a = \bar{a}$ we sometimes will have to suffer with d_a real and negative. See section 10.4.

¹³We have already seen examples where $d_a < 0$ (see exercise 2.2). In such a case we instead get $d_a = |d_a|$.

¹⁴Once again if $N_{ab}^c > 1$ there are additional indices at the vertices and these must match as well. See section 10.6.1.

$$\begin{array}{c} \uparrow \\ a \end{array} \begin{array}{c} \uparrow \\ b \end{array} = \sum_c \sqrt{\frac{d_c}{d_a d_b}} \begin{array}{c} \nearrow a \\ \searrow b \\ \uparrow c \end{array}$$

Fig. 10.21 Insertion of a complete set of states with isotopy invariant normalization of diagrams. See Fig. 11.4 for how to interpret the square root in cases where $d < 0$.

$$\begin{array}{c} \uparrow d \\ \nearrow a \\ \searrow b \\ \uparrow c \end{array} = \delta_{cd} \sqrt{\frac{d_a d_b}{d_c}} \begin{array}{c} \uparrow \\ c \end{array}$$

Fig. 10.22 Bubble diagram with isotopy invariant normalization of diagrams. See Fig. 11.2 for how to interpret the square root in cases where $d < 0$.

¹⁵Indeed, the reason why we changed the value of all vertices, as in Fig. 10.20, and not just rescale the vertex corresponding to a simple loop as in Fig. 10.18, is in order to keep F from changing.

A crucial point is that the F -matrix does not need any alteration when we switch from physics to isotopy-invariant normalization¹⁵! One can check that in changing normalizations both sides of Fig. 9.1 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 diagrams are exactly the same as those described in section 10.2.2 except that loops are now normalized with the quantum dimension as in Fig. 10.18 and our orthonormality relationships (Fig. 10.9 and Fig. 10.10) are altered to those shown in Fig. 10.21 and Fig. 10.22.

10.3.1 Further 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. 10.16 (up to a possible sign, which we will discuss in section 10.4). However, we emphasize this does not guarantee full 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. 10.13 cannot generically be turned into an equality. In Fig. 10.23 and 10.24 we give similar examples and show why we should not expect the two pictures on the far right and far left of these figures to evaluate to the same result.

$$\begin{array}{c} \nearrow c \\ \searrow \bar{c} \\ \uparrow a \end{array} = [F_a^{c\bar{c}a}]_{Ib} \begin{array}{c} \nearrow c \\ \searrow \bar{c} \\ \uparrow a \end{array} = \sqrt{\frac{d_a d_c}{d_b}} [F_a^{c\bar{c}a}]_{Ib} \begin{array}{c} \nearrow c \\ \searrow b \\ \uparrow a \end{array}$$

Fig. 10.23 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. 10.22. Note that if we were to use the physics normalization, the prefactor of $\sqrt{d_b / (d_a d_c)}$ would be absent. Generally we should not expect that the prefactors of d 's and F obtained on the right should cancel each other. However, in chapter 11 we focus on precisely the theories where this prefactor does turn out to cancel as is required for full isotopy invariance.

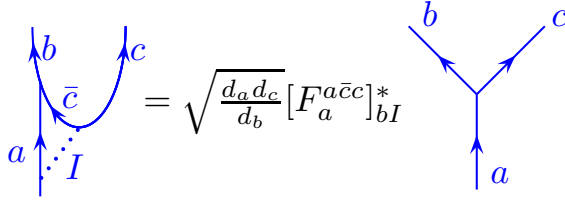


Fig. 10.24 The mirror image of Fig. 10.23. Here we use the fact that F is hermitian, so $F^{-1} = [F^*]^T$.

Thus it seems that our most general theory with causal isotopy invariance cannot achieve full 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. 10.23: 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 the next chapter, there are many theories where one does have full isotopy invariance, and the prefactor incurred in the process shown in Fig. 10.23 turns out to be unity.

¹⁶Having a loop which evaluates to a non-positive number gives a situation where our inner product in Eq. 10.18 is not positive definite, and this can cause all sorts of problems in quantum mechanics (existence of null vectors etc), which we would really like to avoid. While one can have consistent diagrammatic algebras with non-positive-definite inner products, we don't want them to describe quantum mechanics.

10.4 Frobenius-Schur Indicator

Let us return now to the zig-zag Fig. 10.16 and let us be more careful about how we assure that this identity holds. To arrange this, we chose a new normalization for our kets such that the factor $d_a[F_a^{a\bar{a}a}]_{II}$ in Fig. 10.19 is just the identity. However, if $[F_a^{a\bar{a}a}]_{II}$ is not a positive real number this can cause serious problems¹⁶. Fortunately, with a gauge choice, we can fix the phase of $[F_a^{a\bar{a}a}]_{II}$ any way we like unless $a = \bar{a}$. Let us see how this can be done.

On the far left of Fig. 10.19 we have a vertex $|\bar{a}a\rangle$ as well as a vertex which we write as $\langle a\bar{a}|$ (compare to Fig. 10.25). Note that these two vertices are *not* generally hermitian conjugates of each other. By making separate gauge transforms on these two states, these kets can be redefined by an arbitrary phase as discussed in section 9.1.3, and this phase then ends up in $[F_a^{a\bar{a}a}]_{II}$ (See the transformation in Eq. 9.5). Thus by a gauge choice we can choose any phase for $[F_a^{a\bar{a}a}]_{II}$, as long as $a \neq \bar{a}$. However, if $a = \bar{a}$, this scheme fails. 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 10.7 for a three line proof). In the literature one defines a gauge-invariant quantity known as the “Frobenius-Schur indicator” for self-dual particles

$$\kappa_a = \text{sign}[F_a^{aaa}]_{II} \quad (10.3)$$

If the Frobenius-Schur indicator for all particles is positive, our redefini-

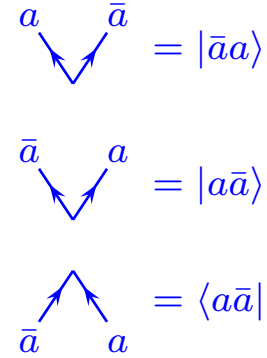


Fig. 10.25 The vertex $|a\bar{a}\rangle$ (top) and the vertex $|\bar{a}a\rangle$ (middle) can be assigned different phases as a gauge choice (See section 9.1.3). The bottom figure here is the hermitian conjugate of the middle and must have the conjugate phase choice. In Fig. 10.19 the leftmost figure includes $|a\bar{a}\rangle$ and $\langle\bar{a}a|$, whereas the phases cancel in the loop formed in the middle picture of Fig. 10.19 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 9.1.3 for discussion of the effects of gauge transform on F).

tion of the normalization of the kets (Figs. 10.18 and 10.20) successfully allows the zig-zag identity to hold as shown in Fig. 10.19. However, if the Frobenius-Schur a minus sign is incurred every time we straighten a zig-zag (as in Fig. 10.16) and no gauge choice can remove this sign.

There are several different approaches for keeping track of the minus signs incurred when there are negative Frobenius-Schur indicators. An approach used by several mathematicians (Turaev [1994]; 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. A similar approach is used by Lin and Levin [2014]. The approach used by Kitaev [2006] and Bonderson [2007] labels creations and annihilations with flags that represent signs. In section 11.3 we will explain a simple but fairly useful method for bookkeeping of Frobenius-Schur indicators which, while not completely general, works for most cases of interest.

10.4.1 Spin 1/2 Analogy

It may seem a bit odd that wiggling a space-time line (as in Fig. 10.16) 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. 9.1 — in this case where a, b, c and e are all labeled with $J = 1/2$. In Fig. 9.1 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 $[F^{\frac{1}{2}\frac{1}{2}\frac{1}{2}}]_{df}$ 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 $[F^{\frac{1}{2}\frac{1}{2}\frac{1}{2}}]_{00}$ 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 two particles fuse to the identity (or singlet $J = d = 0$) can then be written explicitly as

$$|\psi\rangle = \frac{1}{\sqrt{2}} (|\uparrow_1\downarrow_2\rangle - |\downarrow_1\uparrow_2\rangle) \otimes |\uparrow_3\rangle \quad (10.4)$$

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. 10.17.

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. 9.1. We can write the state where the right two fuse to $J = f = 0$ analogously

as

$$|\psi'\rangle = |\uparrow_1\rangle \otimes (|\uparrow_2\downarrow_3\rangle - |\downarrow_2\uparrow_3\rangle) \frac{1}{\sqrt{2}} \quad (10.5)$$

which is precisely analogous to (but the hermitian conjugate of) the top half (the “bra”) of the left hand side of Fig. 10.17.

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.10.17 is¹⁷

$$\langle\psi'|\psi\rangle = -1/2$$

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.

¹⁷This result of $-1/2$ is precisely the $6j$ symbol

$$\left\{ \begin{array}{ccc} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \end{array} \right\}$$

10.5 Appendix: Further Simplifying Assumptions

10.5.1 Pivotal Assumption

A property that may seem obvious is known as the pivotal property. This states that there should be isomorphisms between a vertex with a downturned line and that with an upturned line, such as that shown in Fig. 10.26. While this seems like a rather minor assumption it turns out to be quite powerful, implying

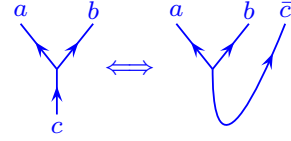


Fig. 10.26 A theory is pivotal if there exists an isomorphism between these two types of vertices.

10.5.2 Spherical Assumption

One typically assumes topological theories are “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. 10.27. The right trace is defined similarly, except that the loop goes to the right of the diagram X as in the right of Fig. 10.27. 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. 10.28. However, the spherical assumption is actually stronger than Fig. 10.28 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. 10.27 The Spherical Property sets the left trace equal to the right trace as shown in the picture.

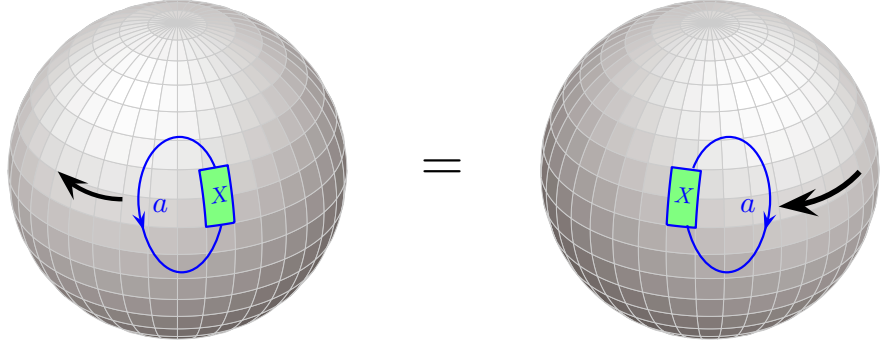


Fig. 10.28 The naming “spherical” comes from the idea that we can pull the string around the back of a sphere (as indicated by the black arrows) to turn a left trace into a right trace.

An obvious result of the spherical assumption is that $d_a = d_{\bar{a}}$. It has been conjectured that any pivotal category is spherical.

10.6 Appendix: Further Details

10.6.1 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. In this case we need to modify Figs. 10.5 and Fig. 10.6. These would instead read

$$\left\langle \begin{array}{c} a' \quad b' \quad c' \quad d' \\ \mu' \quad \nu' \\ e' \quad f' \end{array} \middle| \begin{array}{c} a \quad b \quad c \quad d \\ \mu \quad \nu \\ e \quad f \end{array} \right\rangle = \delta_{aa'} \delta_{bb'} \delta_{cc'} \delta_{dd'} \delta_{ee'} \delta_{ff'} \delta_{f\bar{d}} \delta_{\mu\mu'} \delta_{\nu\nu'}$$

Fig. 10.29 Orthonormality of trees in the case of higher fusion multiplicities. Note that this diagram uses the physics normalization of section 10.2. Here vertices are given indices $\mu \in N_{ab}^e$ and $\nu \in N_{ec}^f$. If we use isotopy invariant normalization the right hand side is multiplied by $\sqrt{d_a d_b d_c d_f}$.

$$\begin{array}{c} | \\ | \end{array} \begin{array}{c} a \\ b \end{array} = \sum_{c, \mu} \begin{array}{c} a \quad \mu \quad b \\ \diagdown \quad \diagup \\ c \\ \diagup \quad \diagdown \\ a \quad \mu \quad b \end{array}$$

Fig. 10.30 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. In the isotopy invariant normalization there is an extra factor of $\sqrt{d_c/(d_a d_b)}$ on the right hand side.

$$\text{Bubble diagram} = \delta_{cd} \delta_{\mu\nu} \text{vertical line } c$$

Fig. 10.31 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. In the isotopy invariant normalization there is an extra factor of $\sqrt{d_a d_b / d_c}$ on the right hand side.

10.7 $[F_a^{aaa}]_{II}$ is real

Let a be a self-dual particle (i.e., $a = \bar{a}$). Working with the physics normalization we already showed (Fig. 10.17) that

$$a \text{ cup} = [F_a^{aaa}]_{II} \text{vertical line } a$$

Similarly using an inverse F -move, and the fact that F is unitary (See section 9.1.1) we can derive

$$a \text{ cap} = [F_a^{aaa}]_{II}^* \text{vertical line } a$$

Assuming only causal isotopy invariance, the equality

$$[F_a^{aaa}]_{II} \text{circle } a = a \text{ cup cap} = \text{cap cup } a = [F_a^{aaa}]_{II}^* \text{circle } a$$

then shows that $[F_a^{aaa}]_{II}$ must be real.

Planar Diagrams with Full 2D Isotopy

11

In the previous chapter, chapter 10, we viewed diagrams as states (kets) of quantum mechanical systems. There was a privileged direction which we called “time” and in general one has to be careful not to change the “causality” of a diagram. While we could make certain deformations of a diagram, not all deformations were generally allowed. In the current chapter, we want to focus on theories that allow a greater degree of topological invariance — theories that enjoy *full isotopy invariance* as shown in Fig. 10.11. That is, any deformation is allowed to a diagram as long as no lines are cut. In this chapter we continue to consider only planar diagrams, so we do not allow over- and under-crossings (which we introduce later in chapter ***). Because of our specialization to these fully isotopy invariant theories, our rules for diagrammatic manipulation will be slightly easier than those in chapter 10.

As in chapter 10 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 ket¹. We can also roughly think of these diagrams as being world lines of particles moving in 1+1 dimensions.

¹We can think of any direction as being time, although it is sometimes most convenient to think of time as up.

In section 11.3 below we will generalize some of our manipulations to also consider a very common modification of our diagrammatic algebra that allows the value of a loop d_a to be negative. Due to this negative value, the theory naively appears to not be valid for describing unitary quantum mechanical system. However, with a minor reinterpretation of our diagrams, these theories can describe unitary quantum mechanical systems but for particles having negative Frobenius-Schur indicators. Such diagrammatic algebras are quite common and include many simple topological theories like $SU(2)_k$ Chern-Simons theory.

11.1 Diagrammatic Rules

As in the previous chapter we would like to use F -matrices to help us convert one diagram into another. We found that bending lines up and down as in Fig. 10.23 can incur nontrivial factors. Here, however, we instead assume that such factors are unity and we can turn up and down the branches of the diagrams freely, thus simplifying manipulations. 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]. Our F -matrix will relate diagrams as shown in Fig. 11.1.

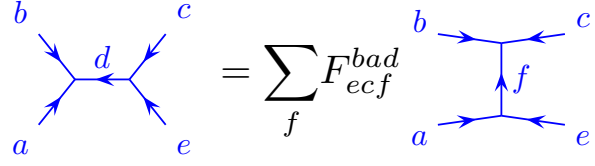


Fig. 11.1 The definition of the F -matrix for 2D isotopy invariant planar diagrams. These are the conventions of Levin and Wen [2005]. For a unitary theory the F -matrix with fixed indices a, b, c, d is unitary in the indices d and f . Our convention is that for this F matrix to be nonzero, the vertices in the pictures must be allowed fusions — i.e., $N_{abd} = N_{ce\bar{d}} = N_{bcf} = N_{ae\bar{f}} = 1$. The case with fusion multiplicities N greater than one is considered in section 11.5.1.

In this chapter, the orientation of this diagram (how we direct the legs compared to some direction we call time) does not matter. We can compare the definition of F -matrix in Fig. 11.1 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 bending legs in Fig. 9.1 and reversing arrows to make it look like Fig. 11.1 and we thereby derive the relation between the two definitions

$$F_{ecf}^{bad} = [F_e^{\bar{a}\bar{b}\bar{c}}]_{df} \quad (11.1)$$

Again the idea of the F -matrix is to write a single diagram (on the left of Eq. 11.1) 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, the contraction of a bubble as shown in Fig. 11.2

Fig. 11.2 The locality principle for the isotopy invariant diagrammatic algebra. 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. 10.22 only written sideways (in this chapter direction on the page does not matter). In particular this rule implies the “no-tadpole” rule, that any diagram of the sort shown in Fig. 11.3 must vanish unless the incoming line is the vacuum.

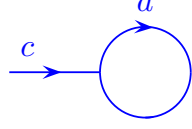


Fig. 11.3 Picture of a tadpole. (Apparently this picture is supposed to look like a tadpole.). The locality principle Fig. 11.2 implies that any diagram containing a tadpole must vanish unless the incoming line is labeled with the vacuum. (I.e., unless there is no incoming line!). Famously, Physical 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. 11.4.

$$\begin{array}{c} \uparrow \\ a \end{array} \begin{array}{c} \uparrow \\ b \end{array} = \sum_c \sqrt{\frac{d_c}{d_a d_b}} \begin{array}{c} a \quad b \\ \diagdown \quad \diagup \\ \quad c \\ \diagup \quad \diagdown \\ a \quad b \end{array}$$

Fig. 11.4 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. 10.21, only now we can orient the diagram in any direction.

The last rule we need is to give a value to the a labelled loop as in Fig. 11.5. 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 and the definition given in Eq. 3.8. In chapter 12 we will show that these two definitions are in fact the same!

$$\begin{array}{c} \circlearrowright \\ a \end{array} = d_a = d_{\bar{a}}$$

Fig. 11.5 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, and we will always assume d_a to be positive. As emphasized in section 10.3, giving the loop this normalization implies we are working with non-normalized kets. (See Fig. 10.18, and also note 18 of chapter 2.)

It is entertaining to note that the value of the loop in Fig. 11.5 is actually contained in the information given in Fig. 11.2 by setting $a = d = I$, which then requires $c = \bar{b}$ and we use $d_c = d_{\bar{c}}$.

²Any planar diagram with no loose ends. As described in detail in 10.1 a diagram with loose ends should be considered a bra or ket or operator.

³We assume in this section that for self-dual particles all the Frobenius-Schur indicators (see section 10.4) are positive so that we need not worry about minus signs when we straighten a zig-zag diagram as in Fig. 10.16.

11.1.1 Summary of Planar Diagram Rules For Full Isotopy Invariant Theories

Given the F -moves (Fig. 11.1) and the locality principle (Fig. 11.2) and the value of the loop d_a , we can evaluate any planar diagram² and turn it into a 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 full isotopy invariant theories. These rules are analogous to those presented in section 10.2.2, only here the rules are simpler³

- (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. 10.14.
- (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).
- (6) One can use F -moves to change the structure of diagrams.
- (7) One can use relations Fig. 11.4 and 11.2 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 .

11.2 Constraints and Examples

There are many constraints on our diagrammatic algebras for 2D 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. 9.3 can be converted to the notation of this chapter (See note ?? of this chapter) to give⁴

$$F_{edl}^{c\bar{f}g} F_{e\bar{l}k}^{baf} = \sum_h F_{gch}^{baf} F_{edk}^{\bar{h}ag} F_{kdl}^{cbh} \quad (11.2)$$

⁴In deriving Eq. 11.2 from Eq. 9.3 we have taken $a, b, c, d \rightarrow \bar{a}, \bar{b}, \bar{c}, \bar{d}$ for ease of notation.

Constraint: Relating F to d

For any theory with full 2D-isotopy (or “almost” 2D-isotopy as we have been considering in this chapter), the value of d_a should be fixed by the

F -matrices:

$$d_a = \frac{1}{F_{aaI}^{\bar{a}aI}} \quad (11.3)$$

This is demonstrated by the manipulations of Fig. 10.19, converted into the notation of the current chapter. In this section we assume that $F_{aaI}^{\bar{a}aI} > 0$. We consider case where it is negative in section 11.3 below.

Constraint: Inversion

One can perform an F -move on the right hand side of Fig. 11.1 to bring it back into the form on the left. We obtain the diagrammatic relation shown in Fig. 11.6,

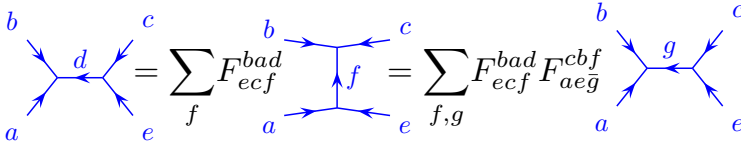


Fig. 11.6 In the second step we apply the same F -matrix equation from Fig. 11.1, but the diagram is rotated by 90 degrees.

which necessarily implies the consistency condition

$$\sum_f F_{ecf}^{bad} F_{aeg}^{cbf} = \delta_{dg} \quad (11.4)$$

Constraint: Rotation

Rotating the diagram in Fig. 11.1 by 180 degrees and comparing it to the original diagram, one derives

$$F_{ecf}^{bad} = F_{ba\bar{f}}^{ec\bar{d}} \quad (11.5)$$

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. 10.23. 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

$$F_{a\bar{a}b}^{c\bar{c}I} = \sqrt{\frac{d_b}{d_a d_c}} \quad (11.6)$$

whenever $b \times c = a + \dots$

Constraint: Unitarity

As mentioned above in Fig. 11.1, the F -matrix, being a change of basis, must be unitary⁵. This means that

$$\sum_f F_{ecf}^{bad} [F_{ecf}^{bad'}]^* = \delta_{dd'} \quad (11.7)$$

$$\sum_d F_{ecf}^{bad} [F_{ecf'}^{bad}]^* = \delta_{ff'} \quad (11.8)$$

⁵For a proper quantum mechanical interpretation we will insist on unitarity. However, planar diagrammatic algebras can be fully self-consistent and fully isotopic invariant without satisfying this unitarity condition. We will discuss an example of this in chapter ??.

or equivalently $[F_{ec}^{ba}]^\dagger = [F_{ec}^{ba}]^{-1}$. Comparing the former to Eq. 11.4 we obtain

$$[F_{ecf}^{bad}]^* = F_{aed}^{cbf} \quad (11.9)$$

Constraint: Hermitian Conjugate

Using reflection across the horizontal axis as in Fig. 10.3, we can reflect the F -matrix equation Fig. 11.1 and compare the reflected to the unreflected diagram to obtain

$$F_{ecf}^{bad} = [F_{\bar{c}\bar{e}\bar{f}}^{\bar{a}\bar{b}\bar{d}}]^* \quad (11.10)$$

$$= [F_{\bar{a}\bar{b}\bar{f}}^{\bar{c}\bar{e}\bar{d}}]^* \quad (11.11)$$

$$= F_{\bar{b}\bar{c}\bar{d}}^{\bar{e}\bar{a}\bar{f}}, \quad (11.12)$$

where in the second line the first line has been used in combination with Eq. 11.5, whereas in the third line the first line has been used in combination with Eq. 11.9.

Constraint: Reflection

An independent condition that is very often imposed is that it should be invariant under left-right reflection. Compare the diagram shown in Fig. 11.7 to that of Fig. 11.1.

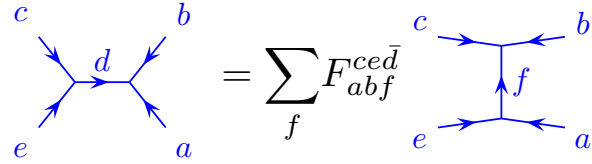


Fig. 11.7 The diagrammatic equation in Fig. 11.1 after being left-right reflected.

If a theory has left-right reflection symmetry, then we must have a further constraint

$$F_{ecf}^{bad} = F_{abf}^{ced\bar{d}} \quad (11.13)$$

While this additional condition is not required for a diagrammatic algebra, and one can even have full isotopy invariance in two dimensions without it, it is often assumed. For three dimensional theories, such a symmetry is natural.

Using Eq. 11.13 along with Eq. 11.12 gives us the natural seeming constraint

$$F_{ecf}^{bad} = [F_{\bar{e}\bar{c}\bar{f}}^{\bar{b}\bar{a}\bar{d}}]^* \quad (11.14)$$

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. 11.8.

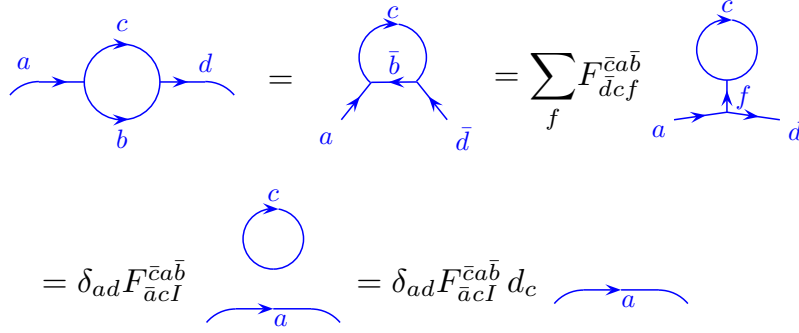


Fig. 11.8 Evaluation of a bubble diagram. In the first step, as usual we can flip the direction of an arrow and turn a particle into its antiparticle. In the second step we apply an F -move (compare to Fig. 11.1). Then by the no-tadpole (locality) rule (Fig. 11.3), we can set f to the vacuum particle I and hence $a = d$.

However, we also know the value of the diagram in Fig. 11.8 from Fig. 11.2 which gives us $\sqrt{d_c d_b / d_a}$. Thus we derive $F_{acI}^{c a b} d_c = \sqrt{\frac{d_c d_b}{d_a}}$, or equivalently (while replacing b with \bar{b} for simplicity and using $d_b = d_{\bar{b}}$) we have

$$F_{acI}^{c a b} = \sqrt{\frac{d_b}{d_a d_c}} \quad (11.15)$$

whenever $c \times b = a + \dots$. Note that Eq. 11.15 could also be obtained from Eq. 11.6 with Eq. ??.

Example: The Theta diagram

A commonly considered diagram is the theta diagram $\Theta(a, b, c)$ shown in Fig. 11.9. This diagram is easily evaluated by using Fig. 11.2 along with the value of a single bubble Fig. 11.5.

$$\Theta(a, b, c) = \text{diagram} = \text{diagram} = d_c \sqrt{\frac{d_a d_b}{d_c}}$$

Fig. 11.9 The Theta diagram. This is evaluated by using Fig. 11.2 along with the value of a single bubble Fig. 11.5.

Example: The tetrahedral diagram

Let us consider one more evaluation known as the tetrahedral diagram as shown in Fig. 11.10. 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.

$$\begin{aligned}
& \text{tetrahedron}(f, b, c, a, d, e) = \text{figure-eight}(b, c, d, a, e) = \sum_g F_{ecg}^{bad} \text{tetrahedron}(f, b, c, g, a, e) \\
& = F_{ecf}^{bad} d_f \sqrt{\frac{d_b d_c}{d_f}} \sqrt{\frac{d_a d_e}{d_f}} \equiv G_{ecf}^{bad}
\end{aligned}$$

Fig. 11.10 Evaluation of the tetrahedral diagram. The first step is just smooth deformation. The second step is application of an F move. Using Fig. 11.2, 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 .

For isotopy invariant theories, the tetrahedral diagram has some obvious symmetries. For example, we should have rotational symmetry in the plane as shown in Fig. 11.11

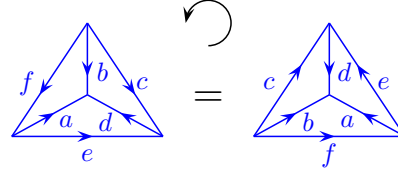


Fig. 11.11 An obvious rotational symmetry of the tetrahedral diagram.

which implies the identity (note the definition of G in Fig. 11.10)

$$G_{ecf}^{bad} = G_{f\bar{e}\bar{c}}^{dba} \quad . \quad (11.16)$$

Another symmetry comes from Eq. 11.5

$$G_{ecf}^{bad} = G_{ba\bar{f}}^{ecd} \quad . \quad (11.17)$$

which we draw as shown in Fig. 11.12.

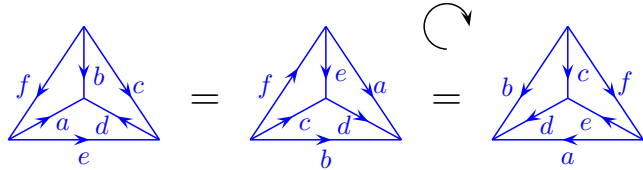


Fig. 11.12 The first step is the identity in Eq. 11.17 and the second step is a rotation as in Fig. 11.11. Although this is actually a planar diagram it appears as a rotation in 3D.

Although the diagram shown in Fig. 11.12 is a planar diagram, from the far left to the far right, it appears as if it is a rotation in 3D. Using Fig. 11.11 and 11.12 we can rotate this tetrahedron in any way we like. If one assumes the reflection symmetry Eq. 11.13 then one can also take the mirror image of the tetrahedron as well to obtain an equivalence between 24 tetrahedral diagrams related by symmetries⁶.

⁶For an example of a spherical category that cannot be put in a form with full tetrahedral symmetry, see Hong [2009].

11.3 Unitarity, Negative d_a , and Frobenius-Schur, again

One often encounters diagrammatic algebras (with evaluation rules as in section 11.1.1) that have full isotopy invariance, but also have a negative value for the variable d_a which gives the value of a loop as in Fig. 11.5. Having such a negative d_a corresponds to a negative normed state (interpreting a loop as the norm of a state as in Fig. 10.18), which implies non-unitarity, which is forbidden in quantum mechanics. Fortunately, it is very often possible to *reinterpret* such diagrammatic algebras as well-behaved unitary quantum theories, having particles with positive-definite inner product but negative Frobenius-Schur indicators (See section 10.4). Because this situation is so common⁷, we will describe it here in some detail.

First, we must elaborate the conditions for this unitary reinterpretation to be possible. We will always assume that

$$\text{sign}(d_a) \text{sign}(d_b) = \text{sign}(d_c) \quad \text{whenever} \quad a \times b = c + \dots \quad (11.18)$$

If this relationship holds (and it very often does!) we can go ahead with our unitary reinterpretation of the diagrammatic algebra.

As a quick aside, we remind the reader that we have suggested a number of times that d_a will be related to the quantum dimension d_a that tells us how the size of the Hilbert space grows with increasing number of particles (See Eq. 3.8). While d_a is here allowed to be negative, d_a must be positive. The general relation will be

$$d_a = |d_a| \quad (11.19)$$

which we will prove in chapter 12.

Before continuing we should probably clarify some issues that may arise in theories with negative d . In some of the diagrammatic calculus we run into factors of \sqrt{d} (See, for example, Figs. 11.2, 11.4, 11.9, 11.10), and this may appear ambiguous. Note, however, that due to Eq. 11.18, the argument of the square root is always positive, so we need only decide on whether the square root should be interpreted as positive (as we have been implicitly doing so far) or negative. The answer to this is written in the captions of Figs. 11.4 and 11.2 — in other words the square root should be interpreted negative⁸ when the factor $(d_a d_b)/d_c$ has both $d_a < 0$ and $d_b < 0$.

Assuming we now have a well defined diagrammatic algebra, with some negative d 's satisfying Eq. 11.18, we now explain how this algebra can be reinterpreted as a well defined unitary quantum mechanical theory. Let us define

$$\kappa_a = \text{sign}(d_a)$$

which we will call the Frobenius-Schur indicator⁹ for particle type a . To remind the reader, when we introduced Frobenius-Schur indicator in section 10.4, a negative κ_a means that if we pull straight a zig-zag in space time (as shown in Fig. 11.13) we incur a minus sign.

⁷Indeed, we have already run into negative d in exercise 2.2 where $d = -1$, and we will re-examine this case in chapter 15 below as well as in ****.

⁸To see that this makes sense, consider Fig. 11.2 in the case where the incoming lines are the identity, but $a = \bar{b}$ and $d_a < 0$.

⁹The common definition of Frobenius-Schur indicator applies only to particles which are self-dual $a = \bar{a}$ as given in Eq. 10.3, or possibly even assigns a value of 0 to non-self dual particles. Here we will assign $\kappa = \pm 1$ to each particle.



Fig. 11.13 A zig-zag in space time. If time in this picture is pointing up, and if the particle a has negative Frobenius-Schur indicator, it incurs a minus sign when the zig-zag is pulled straight.

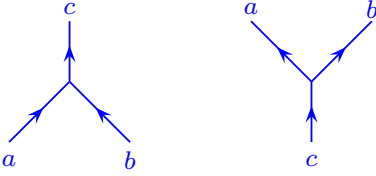


Fig. 11.14 With time going vertical, the left diagram is a Frobenius-Schur cap if and only if $\kappa_a = \kappa_b = -1$. (The directions of the arrows do not matter, and if the particles are self-dual we do not draw arrows). The right diagram is never a Frobenius-Schur cap.



Fig. 11.15 With time going vertical, the left diagram is a Frobenius-Schur cap if and only if $\kappa_a = -1$. The right diagram is never a Frobenius-Schur cap. We can think of these diagrams as being the same as the diagrams in Fig. 11.14 with c being the identity. The directions of the arrows do not matter.

Our reinterpretation of this diagrammatic algebra is now quite simple. We evaluate isotopy invariant diagrams using the rules given in section 11.1.1. We call this result the *non-unitary* evaluation of the diagram as it corresponds to the non-unitary theory with negative d 's. However, we now insert two additional crucial rules into our list

- (0.a) We break the space-time symmetry and declare one direction on the page (usually up) to represent time.
- (0.b) Before evaluating a diagram, count the number of Frobenius-Schur caps, and call it n . After fully evaluating the diagram multiply the final result by $(-1)^n$.

Here a Frobenius-Schur cap occurs when we go forward in time and two particles with $\kappa = -1$ come together to annihilate or form a particle having $\kappa = +1$. (See examples in Figs. 11.14 and 11.15). Another way of counting the Frobenius-Schur caps is to imagine erasing all lines in the diagram which have $\kappa = +1$. This leaves only a set of closed loops (due to Eq. 11.18). We then just need to count caps in this set of closed loop of the form shown in the left of Fig. 11.15.

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. As a simple example, consider the evaluation of a single loop as in Fig. 11.5 where $d_a < 0$. In this case we assign $\kappa_a < 0$ for this particle type. Before evaluating the loop we count that there is a single Frobenius-Schur cap on the top of the loop (as in the left of Fig. 11.15). We evaluate the diagram with the rules of section 11.1.1, to obtain $d_a < 0$ as the nonunitary evaluation. However, applying rule (0.b) 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. 10.18.)

As a second example, consider in the middle right of Fig. 11.9, 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 sign coming from our above discussed rule of how to handle square roots with negative d 's). However, applying rule 0, there is a single Frobenius-Schur 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 positive as it should be for a diagram that can be written as $\langle \text{state} | \text{state} \rangle$ as in Fig. 10.6.

As a third example, consider the same diagram (middle right of Fig. 11.9) 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 Frobenius-Schur 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 is fully isotopy invariant but has negative d_a , which can then be interpreted as

a unitary theory with negative Frobenius-Schur indicators — is quite common. There are many topological theories of this type — including the semion theory $SU(2)_1$ and more generally theories like $SU(2)_k$.

11.4 More General Diagrammatic Algebras and Important Warning

The diagrammatic algebra we have defined in this chapter (See summary *******) make a number of simplifying assumptions — full isotopy invariance, reflection symmetry, unitarity, and so forth. These are common assumptions to make for physical systems. However, particularly in the mathematics literature, one will find discussion of many diagrammatic algebras where some of these simplifications do not hold. The **important warning** is that one should be particularly careful to make sure that you know what assumptions do, or do not, hold for any particular case you happen to be considering.

In the math literature the diagrammatic algebras we have been developing are known as *tensor fusion categories*. There is a very detailed classification of fusion categories describing particular assumptions that one can impose and the resulting properties. We do not have time (or inclination) to make a complete classification here.

11.5 Appendix: Isotopy Invariant F -matrix

Odds and Ends

11.5.1 F -matrix with higher fusion multiplicities

As in section 9.1.2 when fusion multiplicities are greater than one, the vertices have additional indices which we label with greek indices μ, ν, \dots . For example, if a and b fuse to c with $N_{ab}^c > 1$ then the vertex will have an additional index $\mu \in 1 \dots N_{ab}^c$. Note that compared to section 9.1.2 we do not put black dots on the vertices here. In the conventions of the current chapter we would then have

$$\begin{array}{c} b \\ \swarrow \\ \mu \\ \swarrow \quad \searrow \\ a \quad d \quad \nu \\ \swarrow \quad \searrow \\ c \quad e \end{array} = \sum_{f \lambda \tau} [F_{ecf}^{bad}]^{\mu\nu}_{\lambda\tau} \begin{array}{c} b \quad \lambda \quad c \\ \swarrow \quad \uparrow \quad \searrow \\ a \quad f \quad e \\ \swarrow \quad \searrow \\ \tau \end{array}$$

Fig. 11.16 F -matrix with fusion multiplicity.

$$c \xrightarrow{\mu} \text{loop}(a, b) \xrightarrow{\nu} d = \delta_{cd} \delta_{\mu\nu} \sqrt{\frac{d_a d_b}{d_c}} c \xrightarrow{\mu} \text{loop}(a, b) \xrightarrow{\nu} d$$

Fig. 11.17 The locality principle for the isotopy invariant diagrammatic algebra with fusion multiplicity.

$$a \uparrow \quad b \uparrow = \sum_{c, \mu} \sqrt{\frac{d_c}{d_a d_b}} \begin{array}{c} a \searrow \mu \nearrow b \\ \quad \quad \quad c \\ a \nearrow \mu \searrow b \end{array}$$

Fig. 11.18 Insertion of a complete set of states with fusion multiplicity.

11.6 Summary of properties of F

The defining equations (See Yuting Hu, Levin Wen.) Note Yuting Hu Frobenius-Schur indicator factor in later paper.

Quantum Dimension

We have claimed several times that the value of a loop d_a in our diagrammatic calculus (See Fig. 10.18 or 11.5) is related to the quantum dimension \mathbf{d}_a which tells us that the dimension of the Hilbert space grows when we fusing n anyons of type a scales as $\sim \mathbf{d}_a^n$ for large n . (See Eq. 3.8). In Eq. 8.8 we also showed that \mathbf{d}_a is the largest eigenvalue of the fusion multiplicity matrix N_a . To show how d_a is related to \mathbf{d}_a we consider fusing two loops labeled a and b as shown in Fig. 12.1.

$$\begin{aligned}
 \text{Diagram 1} &= \sum_c \sqrt{\frac{d_c}{d_a d_b}} \text{Diagram 2} \\
 &= \sum_c \sqrt{\frac{d_c}{d_a d_b}} \text{Diagram 3} = \sum_c N_{ab}^c \text{Diagram 4}
 \end{aligned}$$

Fig. 12.1 Fusing two loops into a single loop. In the first line we use the completeness relation Fig. 10.21, then we deform to the second line and finally in the last step we remove the bubble using Fig. 10.22.

The result seems rather natural, that a and b can fuse together to form c in all possible ways. The derivation¹ uses the completeness relation in the first line (Fig. 10.21), then we deform to get to the second line, and finally in the last step we remove the bubble using Fig. 10.22.

We have claimed that the value of a loop of string a will be given by $\mathbf{d}_a = |d_a|$ in general². This then gives Fig. 12.1 the algebraic interpretation that the quantities \mathbf{d}_a 's are a representation of the fusion algebra

$$\mathbf{d}_a \mathbf{d}_b = \sum_c N_{ab}^c \mathbf{d}_c \quad (12.1)$$

Now, given Eq. 12.1, which we derived using diagrammatic algebra (and in particular using the fact that a loop evaluates to the quantity $\mathbf{d}_a = |d_a|$), we can show that \mathbf{d}_a is the largest eigenvalue of the fusion multiplicity matrix.

²For the case where $d_a < 0$ here we mean the unitary evaluation which gives $|d_a|$ as discussed in section 11.3

¹The 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 consider additional indices at the vertices, in which case we use Fig. 10.30 and Fig. 10.31 in place of Fig. 10.21 and Fig. 10.22 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 10.5.1).

To do this let us think of the fusion multiplicity for particle a , as a matrix N_a with indices b and c , as we did in Eq. 8.8. We can then think of d_c for the different indices c as a vector \vec{d} , and we can rewrite Eq. 12.1 as

$$d_a \vec{d} = [N_a] \vec{d}$$

I.e, the vector \vec{d} is an eigenvector of N_a with eigenvalue d_a .

Note now that the fusion multiplicity matrix N_a has only non-negative elements and \vec{d} has only positive elements. This allows us to apply the Perron-Frobenius theorem which says that for matrices with only non-negative elements³ there is a unique eigenvector with all positive entries, and it corresponds to the largest eigenvalue. Thus we conclude that d_a is actually the largest eigenvalue of the matrix N_a and it has eigenvector \vec{d} . Thus we conclude that the value of a loop d_a is the same as the definition of the quantum dimension d_a as largest eigenvalue of the fusion multiplicity matrix.

Since d_a and d_a are so closely related one often abuses nomenclature and refers to d_a as being the quantum dimension d_a whereas they need only be the same in absolute value.

³If N_a had all positive entries, the Perron-Frobenius eigenvalue would be the unique largest eigenvalue in absolute value. However, if N_a has entries which are zero, it is possible that there are other eigenvalues with the same absolute value as the Perron-Frobenius eigenvalue.

Some Simple Examples of Planar Diagram Algebras

13

In this chapter we give a few examples of simple planar diagram algebras. To a mathematician these are known as spherical tensor categories (See section 10.5.2). All of the examples given here will enjoy full isotopy invariance (perhaps with a nontrivial Frobenius-Schur indicator in some cases) so we will use the notation of chapter 11.

13.1 \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

$$1 \times 1 = 0$$

which tells us that 1 is its own antiparticle $1 = \bar{1}$. This is known as \mathbb{Z}_2 fusion rules and is shown in Fig. 13.1.

Our fusion multiplicity matrix is $N_{110} = N_{101} = N_{011} = N_{000} = 1$ and $N_{111} = N_{001} = N_{010} = N_{100} = 0$. (Since N is symmetric in its indices, we could have just written $N_{110} = N_{000} = 1$ and $N_{100} = N_{111} = 0$.)

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. 13.2. If one does not draw the identity particle, diagrams must then be just a so-called *loop gas* as shown in Fig. 13.3. The constraint $N_{100} = 0$ means that loops cannot end, and $N_{111} = 0$ means that loops cannot intersect. With these fusion rules,

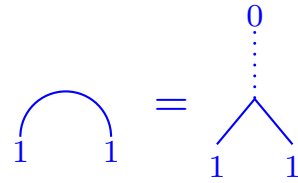


Fig. 13.1 Fusing two 1-particles to the vacuum, shown in two notations.

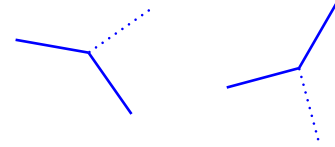


Fig. 13.2 Examples of allowed vertices for the \mathbb{Z}_2 fusion rules. A 1 particle (drawn solid) comes into the vertex and the 1-particle must also go out of the same vertex. The 0 particle, the identity, is drawn dotted, but it need not be drawn at all.



Fig. 13.3 A loop gas has \mathbb{Z}_2 fusion rules.

there are two sets of F -matrices and d 's that give a consistent fully isotopy invariant planar algebras. Note that the identity particle is always assigned quantum dimension $d_0 = 1$.

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 15.1 below.

13.1.1 $d = 1$ loop gas

Here we choose $d_1 = 1$ for the nontrivial particle, in which case every F which is consistent with the fusion rules (See Fig 11.1) is $+1$. In other words, $F_{ecf}^{bad} = 1$ 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

$$F_{000}^{000} = 1/d_0 = 1 \quad (13.1)$$

$$F_{110}^{110} = 1/d_1 = 1 \quad (13.2)$$

$$F_{001}^{110} = F_{111}^{000} = 1 \quad (13.3)$$

$$F_{101}^{101} = F_{011}^{011} = 1 \quad (13.4)$$

$$F_{010}^{101} = F_{100}^{011} = 1 \quad (13.5)$$

The first two lines are required from Eq. 11.3. Eq. 13.3 is from Eq. 11.6. Eq. 13.4 and Eq. 13.5 can be derived from Eq. 13.2 and Eq. 13.3 by the tetrahedral symmetry equation Eq. 11.16.

The $d = 1$ 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.)

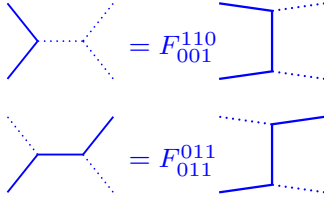


Fig. 13.4 These F -moves for the \mathbb{Z}_2 loop gas simply deform the path of the particles. These are known as “isotopy” moves.

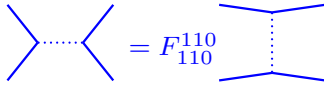


Fig. 13.5 This F -move for the \mathbb{Z}_2 loop gas reconnects the paths of particles. This is known as a “surgery” move.

13.1.2 $d = -1$ 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 . Again $F_{ecf}^{bad} = \pm 1$ for every case where $N_{abd} = N_{ced} = N_{bcf} = N_{aef} = 1$ and is zero otherwise. The signs of F are given as follows

$$F_{000}^{000} = 1/d_0 = 1 \quad (13.6)$$

$$F_{110}^{110} = 1/d_1 = -1 \quad (13.7)$$

$$F_{001}^{110} = F_{111}^{000} = 1 \quad (13.8)$$

$$F_{101}^{101} = F_{011}^{011} = 1 \quad (13.9)$$

$$F_{010}^{101} = F_{100}^{011} = 1 \quad (13.10)$$

As with the $d = 1$ loop gas, the first two lines are required from Eq. 11.3. Eq. 13.3 is from Eq. 11.6. Eq. 13.4 and Eq. 13.5 can be derived from Eq. 13.2 and Eq. 13.3 by the tetrahedral symmetry equation Eq. 11.16. Note in particular how the signs work in Fig. 11.10 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. 11.1 for details of how the F -matrix is

defined). Moves such as shown in Fig. 13.4 simply deform the path of the particle and do not incur a sign. However, the move shown in Fig. 13.5 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

$$\begin{array}{l} \text{Value of } (d = -1) \\ \text{loop diagram} \end{array} = (-1)^{\text{number of loops}}$$

As mentioned in section 11.3, while this is a perfectly consistent diagrammatic algebra, it has non-positive definite inner products and therefore is not appropriate for describing quantum mechanics.

To make a proper unitary theory out of the $d = -1$ loop gas we let there be a Frobenius-Schur indicator

$$\kappa_1 = -1$$

With this nontrivial Frobenius-Schur indicator, and assuming the time direction is up on the page, the value of a diagram is (see section 11.3)

$$\begin{array}{l} \text{Value of } (d = -1) \\ \text{loop diagram} \\ \text{in theory with} \\ \text{nontrivial Frobenius-Schur} \end{array} = (-1)^{\text{number of loops} + \text{number of caps}}$$

For example, in Fig. 13.3 there are 10 loops and 14 caps, so the full value of the diagram is +1.

13.2 Fibonacci Fusion Rules: The Branching Loop Gas

We now consider Fibonacci fusion rules as discussed in sections 8.2.1 and 9.0.1 above. Here the nontrivial fusion rule is¹

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

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. 13.6) so the loop gas can have branches as shown in Fig. 13.7.

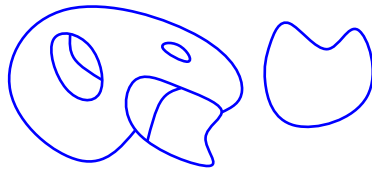


Fig. 13.7 A Fibonacci branching loop diagram allows intersections of loops, but no loop ends.

¹Here we have switched to the notation of τ for the nontrivial particle and I for the vacuum. Using 1 and 0 is also common.

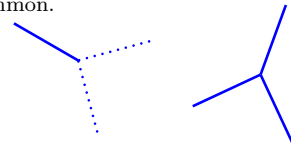


Fig. 13.6 An allowed fusion vertex (right) and a disallowed fusion vertex (left) for the Fibonacci fusion rules. The solid line is τ and the dotted line is the identity. The vertices shown in Fig. 13.2 are also allowed.

The fusion multiplicity matrix N_{abc} is zero if exactly one of the indices is τ and the other two are I . Otherwise $N_{abc} = 1$. As usual F_{ecf}^{bad} is nonzero for every case where $N_{abd} = N_{ced} = N_{bcf} = N_{aef} = 1$ and F is zero otherwise. We can establish the nonzero components of the F -matrices for these fusion rules.

$$F_{III}^{III} = 1/d_I = 1 \quad (13.11)$$

$$F_{\tau\tau I}^{\tau\tau I} = 1/d_\tau \quad (13.12)$$

$$F_{II\tau}^{\tau\tau I} = F_{\tau\tau\tau}^{III} = 1 \quad (13.13)$$

$$F_{\tau\tau\tau}^{\tau\tau I} = 1/\sqrt{d_\tau} \quad (13.14)$$

$$F_{\tau\tau I}^{\tau\tau\tau} = 1/\sqrt{d_\tau} \quad (13.15)$$

$$F_{\tau I\tau}^{\tau I\tau} = F_{I\tau\tau}^{I\tau\tau} = 1 \quad (13.16)$$

$$F_{I\tau I}^{\tau I\tau} = F_{\tau I I}^{I\tau\tau} = 1 \quad (13.17)$$

$$F_{I\tau\tau}^{\tau\tau\tau} = F_{\tau I\tau}^{\tau\tau\tau} = F_{\tau\tau\tau}^{I\tau\tau} = F_{\tau\tau\tau}^{\tau I\tau} = 1 \quad (13.18)$$

$$F_{\tau\tau\tau}^{\tau\tau\tau} = -1/d_\tau \quad (13.19)$$

As with the case of the \mathbb{Z}_2 loop gases, the first two lines are required from Eq. 11.3. Eq. 13.13 and Eq. 13.14 are from Eq. 11.6. Eq. 13.15 comes from Eq. 13.14 and Eq. ???. Eqs. 13.16, 13.17, and 13.18 can be derived from Eqs. 13.12, 13.13, 13.14 and 13.15 by the tetrahedral symmetry equation Eq. 11.16. Finally, Eq. 13.19 comes from the requirement that the two by two matrix $[F_{\tau\tau}^{\tau\tau}]$ is a unitary matrix (See Fig. 11.1) which we write out as

$$F_{\tau\tau}^{\tau\tau} = \begin{pmatrix} 1/d_\tau & 1/\sqrt{d_\tau} \\ 1/\sqrt{d_\tau} & -1/d_\tau \end{pmatrix} \quad (13.20)$$

The unitarity requirement on this matrix also gives us

$$\frac{1}{d_\tau^2} + \frac{1}{d_\tau} = 1 \quad (13.21)$$

The solution to this is

$$d_\tau = \frac{1 + \sqrt{5}}{2}$$

which matches the expected quantum dimension given in Eq. 8.2 as it must, given the considerations of chapter 12. Note that this matrix also matches our previous claim in Eq. 9.2.

Eq. 13.21 also has a solution with $d_\tau < 0$. However, as discussed in section 11.3 this would correspond to a non-unitary theory and in this case it is not possible to remove the problem by using a negative Frobenius-Schur indicator. To see how this fails, recall we can push the minus sign onto the quantum dimension only in cases where $\kappa_a \kappa_b = \kappa_c$ whenever $a \times b = c + \dots$. Here, since $\tau \times \tau = \tau + \dots$ we cannot assign $\kappa_\tau = -1$.

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. 13.4. The nontrivial F -moves (corresponding to the matrix F in Eq. 13.20) are summarized in Fig. 13.8.

$$\begin{aligned}
\text{Diagram 1} &= \frac{1}{d_\tau} \text{Diagram 2} + \sqrt{\frac{1}{d_\tau}} \text{Diagram 3} \\
\text{Diagram 4} &= \sqrt{\frac{1}{d_\tau}} \text{Diagram 5} - \frac{1}{d_\tau} \text{Diagram 6}
\end{aligned}$$

Fig. 13.8 The F -moves for the Fibonacci branching loop gas. Note that the first line is actually the insertion of a complete set of states as in Fig. 11.4

13.3 Ising Fusion Rules: A two species loop gas

As discussed in section 8.2.2 the Ising 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. 13.9) and σ (which we draw red loops in Fig. 13.9). The rule of this loop gas is that one may have a vertex with two sigmas and a ψ , which appears as a blue line splitting off from a red loop.

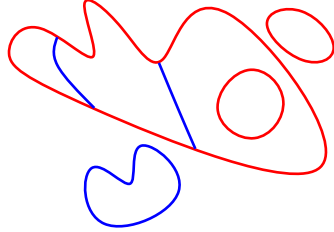


Fig. 13.9 A diagram with Ising fusion rules. Here σ is red and ψ is blue.

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

$$\begin{aligned}
1/d_I &= 1/d_\psi = 1 \\
&= F_{III}^{III} = F_{\psi\psi I}^{\psi\psi I} = F_{II\psi}^{\psi\psi I} = F_{\psi\psi\psi}^{\psi\psi\psi} = F_{\psi I\psi}^{\psi I\psi} = F_{I\psi\psi}^{\psi I\psi} = F_{II\psi}^{\psi I\psi} = F_{\psi II}^{\psi\psi\psi}
\end{aligned}$$

as given in Eqs 13.1-13.5. One might wonder why we do not consider $d_\psi = -1$ with a Frobenius-Schur indicator. This is for the same reason why we could not consider negative Frobenius-Schur indicator in the Fibonacci case. Here we must have $\kappa_\sigma \kappa_\sigma = \kappa_\psi$ so we must have a postive Frobenius-Schur indicator for the ψ particle.

Very similarly we have

$$\begin{aligned} F_{\sigma\sigma I}^{\sigma\sigma I} &= 1/d_\sigma \\ F_{II\sigma}^{\sigma\sigma I} &= F_{\sigma\sigma\sigma}^{III} = 1 \\ F_{\sigma I\sigma}^{\sigma I\sigma} &= F_{I\sigma\sigma}^{I\sigma\sigma} = 1 \\ F_{I\sigma I}^{\sigma I\sigma} &= F_{\sigma II}^{I\sigma\sigma} = 1 \end{aligned}$$

The first equation is from Eq. 11.3, and the second from Eq. 11.6. The last two are derived from the first two via the tetrahedral symmetry Eq. 11.16.

Further using Eqs. 11.6 and 11.15 we obtain

$$F_{\sigma\sigma\psi}^{\sigma\sigma I} = F_{\sigma\sigma I}^{\sigma\sigma\psi} = 1/d_\sigma \quad (13.22)$$

$$F_{\psi\psi\sigma}^{\sigma\sigma I} = F_{\sigma\sigma\sigma}^{\psi\psi I} = F_{\psi\sigma I}^{\sigma\psi\sigma} = F_{\sigma\psi I}^{\psi\sigma\sigma} = 1 \quad (13.23)$$

Enforcing unitarity on the two by two matrix $[F_{\sigma\sigma}^{\sigma\sigma}]$ we get

$$F_{\sigma\sigma\psi}^{\sigma\sigma\psi} = -1/d_\sigma \quad (13.24)$$

giving the two by two matrix the form

$$[F_{\sigma\sigma}^{\sigma\sigma}] = \begin{pmatrix} 1/d_\sigma & 1/d_\sigma \\ 1/d_\sigma & -1/d_\sigma \end{pmatrix} \quad (13.25)$$

The unitarity condition also gives us the condition that

$$d_\sigma = \pm\sqrt{2}$$

which is expected from section 8.2.2. Both of these roots are viable solutions. The positive solution is known as the Ising model whereas the negative solution corresponds to $SU(2)_2$ Chern-Simons theory. As we might expect when we find a negative value of d_σ we must also assign the σ particle a negative Frobenius-Schur indicator $\kappa_\sigma = -1$ so as to keep the theory unitary.

The remaining nonzero elements of F are obtained from Eq. 13.22-13.24 by using tetrahedral symmetry Eq. 11.16 to obtain

$$1 = F_{\sigma\psi\sigma}^{\sigma I\sigma} = F_{\psi\sigma\sigma}^{I\sigma\sigma} = F_{\sigma I\sigma}^{\sigma\psi\sigma} = F_{I\sigma\sigma}^{\psi\sigma\sigma} \quad (13.26)$$

$$= F_{\psi\sigma\psi}^{\sigma I\sigma} = F_{\sigma\psi\psi}^{I\sigma\sigma} = F_{\sigma\sigma\sigma}^{\psi I\psi} = F_{\sigma\sigma\sigma}^{I\psi\psi} \quad (13.27)$$

$$= F_{\sigma I\psi}^{\psi\sigma\sigma} = F_{I\psi\sigma}^{\sigma\sigma\psi} = F_{\psi I\sigma}^{\sigma\sigma\psi} = F_{I\sigma\psi}^{\sigma\psi\sigma} \quad (13.28)$$

$$-1 = F_{\sigma\psi\sigma}^{\sigma\psi\sigma} = F_{\psi\sigma\sigma}^{\psi\sigma\sigma} \quad (13.29)$$

The nontrivial F -moves corresponding to the matrix Eq. 13.25 are shown in Fig. 13.10.

13.4 Rigidity

There are obviously many other diagrammatic algebras we can define (Indeed, we will run into more in chapters *** and *** below). The

$$\begin{aligned}
 \text{Diagram 1} &= \frac{1}{d_\sigma} \text{Diagram 2} + \frac{1}{d_\sigma} \text{Diagram 3} \\
 \text{Diagram 4} &= \frac{1}{d_\sigma} \text{Diagram 5} - \frac{1}{d_\sigma} \text{Diagram 6}
 \end{aligned}$$

Fig. 13.10 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. 11.4

trends we have found here are very general. In particular, as mentioned in section 9.0.2 for any given set of fusion rules, there are a finite number of possible consistent solutions of the pentagon equation thus defining a finite number of possible consistent F -matrices. This property is called “rigidity” Etingof et al. [2015]². This suggests the possibility of building a sort of “periodic table” of all possible diagrammatic algebras by starting with a small number of fields (we chose two fields in the case of the \mathbb{Z}_2 rules and the Fibonacci rules, and then three fields for Ising) then looking for all possible fusion rules with these fields and then finding all possible consistent F -matrices for those fusion rules. While there are an infinite number of possibilities, it is a countable infinite number.

²It is sometimes known as “Ocneanu Rigidity”, having been discovered by Ocneanu although never published by him.

Exercises

Exercise 13.1 Show that evaluation of the diagram in Fig. 13.7 gives $-d_\tau^{9/2}$.

Exercise 13.2 Show that evaluation of the diagram in Fig. 13.9 gives $d_\psi^2 d_\sigma^3 \kappa_\sigma$.

Planar Diagrams from Groups

In this chapter we will use the structure of discrete groups to build rules for planar diagrams. In section 14.1 we will label our diagrams with group elements whereas in section 14.2 we will label diagrams with group representations.

14.1 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. 14.1.

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. 14.2.

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. 14.2 analogous to reversing an arrow in order to turn a particle into its antiparticle.

In cases where the group is abelian so that $g \times h = gh = hg = h \times g$ this structure completely fits with the structure of fusion of particle types we defined in section 8.1 above. In section 14.1.3 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 $d_g = 1$ 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. 14.3.

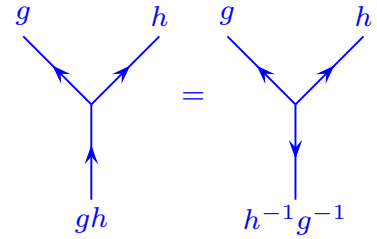


Fig. 14.1 Fusion is defined by group multiplication. On the right we show the three particles oriented as all leaving the vertex. With this orientation when the three particles are multiplied together in clockwise order, they should fuse to the identity $gh(h^{-1}g^{-1}) = h(h^{-1}g^{-1})g = (h^{-1}g^{-1})gh = I$.



Fig. 14.2 Reversing an arrow inverts the group element.

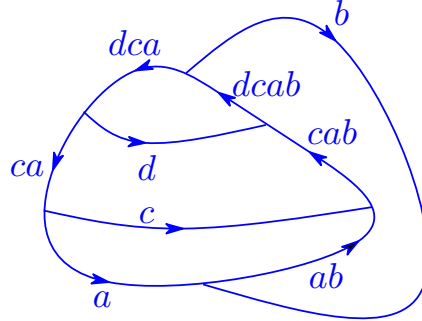


Fig. 14.3 A planar diagram with fusion being defined as group multiplication. For each vertex, if all arrows are pointed out of the vertex, then going around the vertex clockwise, the group elements multiply to the identity, as shown in Fig. 14.1.

14.1.1 Group Cohomology

¹Group 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.

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*.¹

Consider a general group G . A so-called 3-cocycle of the group is given by a function of three variables $\omega(a, b, c)$ where $a, b, c \in G$ that satisfies

$$\omega(a, b, c)\omega(a, bc, d)\omega(b, c, d) = \omega(ab, c, d)\omega(a, b, cd) \quad (14.1)$$

Generally we will consider cases of ω being a $U(1)$ valued complex phase. In group cohomology notation we say that

$$\omega \in H^3(G, U(1)) \quad (14.2)$$

Eq. 14.1 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,

$$[F_{(abc)}^{a,b,c}]_{(ab),(bc)} = \omega(a, b, c)$$

So that we have diagrammatically

Fig. 14.4 The 3-cocycle is precisely an F -matrix. Compare to Fig. 9.1.

Examining the pentagon equation Eq. 9.3 and Fig. 9.5 we see that this is precisely the same as Eq. 14.1 in a different language. Note that

there is no sum over indices here (like the sum over possible elements h in Eq. 9.3) 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 9.1.3). 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. 14.5 to transform the cocycle by

$$\omega(a, b, c) \rightarrow \frac{u(a, bc)u(b, c)}{u(a, b)u(ab, c)} \omega(a, b, c). \quad (14.3)$$

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. 14.1 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. 14.2 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-cocycles. (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 11. 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. 14.6.

Thus for full isotopy invariance (and allowing for d both $+1$ and -1) we need to have

$$s(a, b)\omega(a, a^{-1}, b) = 1 \quad (14.4)$$

$$s(a, b)\omega(a, b^{-1}, b) = 1 \quad (14.5)$$

for all a, b in the group with

$$s(x, y) = \begin{cases} -1 & d_x = d_y = -1 \\ +1 & \text{otherwise} \end{cases} \quad (14.6)$$

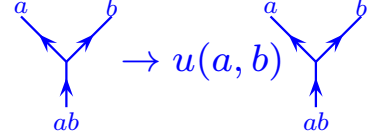


Fig. 14.5 We have the freedom to make a gauge transform of a vertex by multiplying by a phase $u(a, b)$.

While this condition seems quite restrictive, the gauge freedom Eq. 14.3 allows us often to achieve this.

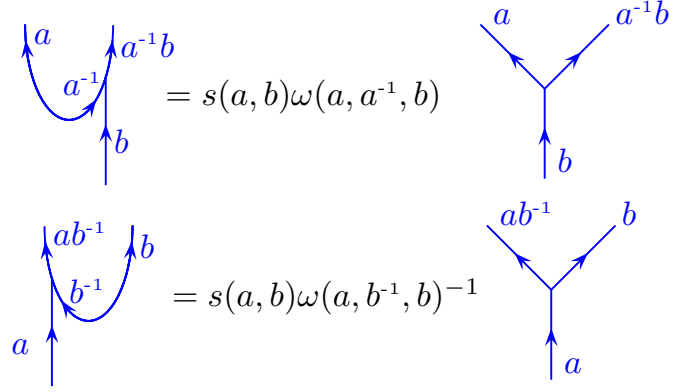


Fig. 14.6 Turning up and down relations (analogous to Fig. 10.23). The prefactor s comes from the proper interpretation of the sign of the \sqrt{d} factors in 10.23. See section 11.3.

A further item to note is that the Frobenius-Schur indicator for a particle is

$$\kappa_a = d_a = \omega(a, a^{-1}, a)$$

and if a is self-dual ($a = a^{-1}$) this is a gauge invariant quantity. If we can do so, we will try to choose a gauge for our 3-cocycles such that we have full isotopy invariance (this is not always possible). In cases where we have negative Frobenius-Schur indicators, we will try to put our 3-cocycles in the form of section 11.3 where we have full isotopy invariance of diagrams up to Frobenius-Schur indicators and correspondingly negative values of d .

14.1.2 Examples: $G = \mathbb{Z}_N$

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 $g \times h = h \times g$ and these particle types fit the description of chapter 8.

The inequivalent 3-cocycles of the group \mathbb{Z}_N can be written as (See references at the end of the chapter)

$$\omega(a, b, c) = \exp \left(\frac{2\pi i p}{N^2} a(b + c - [b + c]_N) \right) \quad (14.7)$$

where here $a, b, c \in 0, \dots, N-1$, and the brackets $[b + c]_N$ means $b + c$ modulo N where the result is chosen to lie in the range $0 \dots N-1$. Here the index p is an integer in the range $0 \dots N-1$ describing the N different gauge-inequivalent 3-cocycles.

The trivial 3-cocycle 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. 14.7, 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 13.1.1.

The only nontrivial 3-cocycle is the $p = 1$ case. Here, using Eq. 14.7 we determine the 3-cocycle is of the form

$$\omega(a, b, c) = \begin{cases} -1 & a = b = c = 1 \\ 1 & \text{otherwise} \end{cases} \quad (14.8)$$

We recognize this as being exactly the case of the $d = -1$ loop gas from section 13.1.2 (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. 13.7, and note that abc here means multiplication with the group operation so is really $(a + b + c) \bmod 2$).

 \mathbb{Z}_3

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. 14.7) and two nontrivial 3-cocycles ($p = 1$ and $p = 2$ in Eq. 14.7). While these nontrivial cocycles provide a valid solution to the pentagon equation 14.1 (or Eq. 9.3) they are not in a form where they enjoy full isotopy invariance. One can use gauge transforms Eq. 14.3 to try to put the cocycles in different forms, but it is not possible to find a gauge where both Eq. 14.4 and Eq. 14.5 are satisfied at the same time². Nonetheless they still provide a consistent planar diagrammatic algebra, although not a fully isotopy invariant one.

 \mathbb{Z}_{odd}

One can further show that \mathbb{Z}_n for any odd n is like the case of \mathbb{Z}_3 . That is, the only fully isotopy invariant case is $p = 0$. See exercise ***.

 \mathbb{Z}_{even}

For \mathbb{Z}_n with n even, the situation is similar to what we just found with \mathbb{Z}_3 for any $p \neq 0, n/2$ (See exercise ***). That is, we have a perfectly consistent planar diagrammatic algebra, although not a fully isotopy invariant one (no matter what gauge we choose).

For the remaining cases $p = 0$ and $p = n/2$, however, we can obtain an isotopy invariant form.

(1) Any even n with $p = 0$

For $p = 0$ we have an isotopy invariant solution given by $\omega(a, b, c) = 1$ for all a, b, c and correspondingly $d = 1$ for all particles³.

²To see that it is not possible to achieve full isotopy invariance note that from Eq. 14.4 and 14.5, 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 ***.

³It is interesting to note that in the case of $n = 4m$, by making a gauge transform we can also express this as an isotopy invariant theory with $d_a = \kappa_a = (-1)^a$ and $\omega(a, b, c) = -1$ when a, b, c are all odd, and $\omega(a, b, c) = 1$ otherwise. This appears similar in spirit to case (2).

(2) $n = 4m + 2$ such that $p = n/2$ is odd

In this case, the particle numbered $2m + 1$ is self dual, and has Frobenius-Schur indicator -1 , which is gauge invariant. A fully isotopy invariant theory is possible by choosing a gauge where $d_a = \kappa_a = (-1)^a$ and $\omega(a, b, c) = -1$ when a, b, c are all odd, and $\omega(a, b, c) = 1$ otherwise. Note that the composition rule Eq. 11.18 is satisfied.

(3) $n = 4m$ such that $p = n/2$ is even

In this case the Frobenius-Schur indicator of the self-dual particle is $+1$ and we can work in a gauge where all $d_a = +1$ and ω is nontrivial

An example of case (3) is that of \mathbb{Z}_4 with $p = 2$. Here we can make a gauge transform (See exercise ***) such that

$$\omega(a, b, c) = \begin{cases} -1 & (a, b, c) = (1, 1, 1); (1, 2, 3); (2, 1, 2); (2, 3, 2); (3, 1, 1) \\ 1 & \text{otherwise} \end{cases} \quad (14.9)$$

and with $d = 1$ for all particles 0,1,2,3. This gives a fully isotopy invariant theory.

14.1.3 Using Nonabelian Groups?

⁴We 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 non-abelian to describe fusion rules where there is more than one fusion channel, such as $g \times h = a + b + \dots$

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 nonabelian⁴ group gh may not be the same as hg .

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, or 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 (as we are in this chapter) 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. 14.1. Thus, at least for planar diagrams we can generalize our rules for particle fusion to allow non-commutative fusions.

All of the figures in this section (Fig. 14.1 – Fig. 14.6) 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.

Example 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 reader⁵ this group has 6 elements which can be written in terms of two generators X and R with multiplication 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 $p = 0, \dots, 5$ in the equation (See references at the end of the chapter)

$$\omega((A, a), (B, b), (C, c)) = \exp\{i\pi p ABC\} \exp\left\{\frac{2\pi i p}{9}(-)^{B+C} a \{(-)^C b + c - [(-)^C b + c]_3\}\right\} \quad (14.10)$$

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 hand side, $\exp(i\pi p ABC)$, matches the two possible 3-cocycles 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. 14.8. 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. 14.10 is trivial. Thus this case of $p = 3$ gives an isotopy invariant cocycle which essentially ignores the a variable of (A, a) and is equivalent to Eq. 14.8 for the A variables with $d_{(A,a)} = (-1)^A$.

14.2 Fusion of Discrete Group Representations

Another way to construct a consistent planar diagrammatic algebra is to work with representations of discrete groups⁶. Suppose we have irreducible representations R_i of a group G . A tensor product of two of these irreducible representations will necessarily decompose into a direct sum of irreducible representations. I.e., we have⁷

$$R_a \otimes R_b \simeq R_c \oplus R_d \oplus \dots \quad (14.11)$$

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 product equation Eq. 14.11 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 $R_{\bar{a}} = R_a^*$.

⁵The group S_3 , the permutation or symmetric group on three elements, represents the symmetries of a triangle and correspondingly is also known as the dihedral group with 6 elements, often denoted D_3 or sometimes D_6 . See section 28.2 for a few more details of this group.

⁶To 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 28.2.4.

⁷If 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. 14.11 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.

⁸Meaning a mapping where the group operation is preserved: $\rho^R(g_1)\rho^R(g_2) = \rho^R(g_1g_2)$.

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 homomorphism⁸ from each group element g to a matrix $\rho_{mn}^R(g)$ (See section 28.2.4). The trace of the representation matrix is known as its character

$$\chi^R(g) = \text{Tr}[\rho^R(g)]$$

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 $\text{Tr}(ab) = \text{Tr}(ba)$ 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

$$\chi^{R_a \oplus R_b}(g) = \chi^{R_a}(g) + \chi^{R_b}(g) \quad (14.12)$$

$$\chi^{R_a \otimes R_b}(g) = \chi^{R_a}(g)\chi^{R_b}(g) \quad (14.13)$$

⁹This orthonormality is derived trivially from the grand orthogonality theorem, Eq. 28.3. 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 .

¹⁰The \oplus 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.

¹¹We have $R_a \otimes R_b \simeq R_b \times R_a$ 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.

Further we have orthonormality relations for irreducible representations:⁹

$$\frac{1}{|G|} \sum_{g \in G} [\chi^{R_a}(g)]^* \chi^{R_b}(g) = \delta_{R_a, R_b} \quad (14.14)$$

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}

$$R_a \otimes R_b \simeq \bigoplus_{c \in \text{irreps}} N_{ab}^c R_c \quad (14.15)$$

where

$$N_{ab}^c = \frac{1}{|G|} \sum_{g \in G} [\chi^{R_c}(g)]^* \chi^{R_a}(g) \chi^{R_b}(g) \quad (14.16)$$

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.)

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 triangle⁵. There are three conjugacy classes, which we will call the identity, the rotations, and the reflections. There are also three irreducible representations¹².

¹²The number of irreducible reps is always equal to the number of conjugacy classes.

	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 14.1 Character table for the group S_3 . Notice the orthogonality of rows as defined by Eq. 14.14.

The group has a character table as given in table 14.1. It is then easy to use Eq. 14.16 to determine the fusion laws for the representations, which are given by

$$I \times I = I, \quad I \times S = S, \quad I \times V = V \quad (14.17)$$

$$S \times S = I, \quad S \times V = V \quad (14.18)$$

$$V \times V = I + S + V \quad (14.19)$$

from which we see that I plays the role of the vacuum particle. Just as an example, let us consider Eq. 14.19. From the character table we have $\chi^V = (2, -1, 0)$ and so $\chi^{V \otimes V} = \chi^V \chi^V = (4, 1, 0) = (1, 1, 1) + (1, 1, -1) + (2, -1, 0) = \chi^I + \chi^S + \chi^V$.

14.2.1 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 methods¹³. In general this can be a bit complicated but the principle is straightforward group theory. As usual we should think of F_{ecf}^{bad} as a basis transform (See Fig. 11.1). 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.

¹³This section uses a bit more advanced group theory and is a bit harder to digest.

- (1) Consider $R_a \otimes R_b \simeq \bigoplus_{\bar{d}} N_{ab}^{\bar{d}} R_{\bar{d}}$ and fuse with $R_c \otimes R_e \simeq \bigoplus_d N_{ce}^d R_d$. The resulting representations, R_d and $R_{\bar{d}}$ then fuse together to form the identity representation. Such a process corresponds to the diagram on the left of Fig. 11.1.
- (2) Consider instead $R_b \otimes R_c \simeq \bigoplus_{\bar{f}} N_{bc}^{\bar{f}} R_{\bar{f}}$ and fuse with $R_a \otimes R_e \simeq \bigoplus_f N_{ae}^f R_f$, and finally fuse $R_{\bar{f}}$ and R_f to form the identity representation. Such a process corresponds to the diagram on the right of Fig. 11.1.

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 F_{ecf}^{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).

To impliment this procedure, we work with explicit D dimensional matrices ρ_{ij}^R for each D dimensional unitary representation R . We extract the identity component of the fusion of the four particles by writing¹⁴

$$\sum_g \rho^{R_a}(g) \otimes \rho^{R_b}(g) \otimes \rho^{R_c}(g) \otimes \rho^{R_e}(g) = C \sum_d \mathbf{w}_d \mathbf{w}_d^\dagger \quad (14.20)$$

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. Similarly we can write

$$\begin{aligned} \sum_g \rho^{R_a}(g) \otimes \rho^{R_c}(g) \otimes \rho^{R_b}(g) \otimes \rho^{R_e}(g) &= C \sum_f \mathbf{z}_f \mathbf{z}_f^\dagger \quad (14.21) \\ &= C \sum_d P \mathbf{w}_d \mathbf{w}_d^\dagger P^T \end{aligned}$$

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. 14.21, 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

$$F_{cef}^{bad} = \mathbf{w}_d^\dagger \cdot \mathbf{z}_f \quad (14.22)$$

The challenge is then simply to extract the correct vectors \mathbf{w}_d and similarly \mathbf{z}_f in Eq. 14.20 and 14.21.

We thus only need to build up the tensor product in Eq. 14.20 and 14.21 step by step. We have¹⁵

$$\rho^{R_a}(g) \otimes \rho^{R_b}(g) = \sum_{\bar{d} \in a \times b} \sum_{\alpha, \beta=1}^{D_{\bar{d}}} \mathbf{x}_{\alpha}^{\bar{d}} [\rho^{R_{\bar{d}}}(g)]_{\alpha\beta} [\mathbf{x}_{\beta}^{\bar{d}}]^\dagger \quad (14.23)$$

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). The particular form of the \mathbf{x} vectors can be extracted using the grand orthogonality theorem Eq. 28.3. Performing the same decomposition for

$$\rho^{R_c}(g) \otimes \rho^{R_e}(g) = \sum_{d \in c \times e} \sum_{\gamma, \delta=1}^{D_d} \mathbf{y}_{\gamma}^d [\rho^{R_d}(g)]_{\gamma\delta} [\mathbf{y}_{\delta}^d]^\dagger \quad (14.24)$$

To find the identity element of the fusion between the tensors in Eq. 14.23 and 14.24 we simply match up the d representations with the \bar{d} representations. Thus we have

$$\begin{aligned} \mathbf{w}_d \mathbf{w}_d^\dagger &\sim \sum_g \sum_{\alpha, \beta, \gamma, \delta=1}^{D_d} \left(\mathbf{x}_{\alpha}^{\bar{d}} [\rho^{R_{\bar{d}}}(g)]_{\alpha\beta} [\mathbf{x}_{\beta}^{\bar{d}}]^\dagger \right) \otimes \left(\mathbf{y}_{\gamma}^d [\rho^{R_d}(g)]_{\gamma\delta} [\mathbf{y}_{\delta}^d]^\dagger \right) \\ &\sim \sum_{\alpha, \beta=1}^{D_d} \left(\mathbf{x}_{\alpha}^{\bar{d}} [\mathbf{x}_{\beta}^{\bar{d}}]^\dagger \right) \otimes \left(\mathbf{y}_{\alpha}^d [\mathbf{y}_{\beta}^d]^\dagger \right) \end{aligned} \quad (14.25)$$

¹⁴This is a result of the grand orthogonality theorem Eq. 28.3. If we sum over all group elements we extract only the identity representation.

¹⁵In cases where $N_{ab}^{\bar{d}} > 1$ we must take extra care to separate to add the multiple instances of each representation. For simplicity let us assume $N_{ab}^{\bar{d}} = 0$ or 1 only.

where in going to the second line we have used the grand orthogonality theorem Eq. 28.3. 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. 14.21, we then determine the F -matrix using Eq. 14.22.

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 *** we walk through calculation of F -matrices for the representations of the group S_3 whose fusion rules we worked out in Eqs. 14.17–14.19 above.

14.2.2 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 28.2.3). 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 angular momentum j can be infinitely 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 16.3 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 representation¹⁶. 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 $j = 0, 1/2, \dots, k/2$ can occur, but one never gets any higher angular momenta.

¹⁶The term “quantum group” is often used. Be warned that a quantum group is not a group.

Further Reading

de Wild, Yuting Hu for group cohomology (probably more)

Yuting Hu Buerschaper for F matrices



Exercises

Exercise 14.1 Some F matrix elements for representations of S_3 [Hard]

Let us consider the simplest nonabelian group S_3 , which we discuss in sections 4, 11, and 28.2.1.

We remind the reader that this group has 6 elements which can be written in terms of two generators X and R with multiplication 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 14.1).

The three representations are as follows: The trivial representation 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 representation in a unitary form as

$$\rho^V(X) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \quad \rho^V(R) = \frac{-1}{2} \begin{pmatrix} 1 & \sqrt{3} \\ -\sqrt{3} & 1 \end{pmatrix}$$

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!)¹⁷

Note that we already know the fusion rules for these representations as they are given in Eqs. 14.17–14.19.

In this exercise we will calculate some F -matrix elements by focusing on the most interesting case, where all four incoming lines in Fig. 11.1 are in the two dimensional V representation. Thus we are interested in the unitary matrix F_{VVf}^{VVd} .

(a) Using the grand orthogonality theorem (Eq. 28.3) find the decomposition

$$\rho^V(g) \otimes \rho^V(g) = \mathbf{x}^I [\mathbf{x}^I]^\dagger + \rho^S(g) \mathbf{x}^S [\mathbf{x}^S]^\dagger + \sum_{\alpha, \beta=1}^2 \mathbf{x}_\alpha^{\bar{V}} [\rho^{R_d}(g)]_{\alpha\beta} [\mathbf{x}_\beta^{\bar{V}}]^\dagger$$

Hint: The one dimensional representations are easy to obtain since they can be obtained by

$$(1/|G|) \sum_g \rho^R(g)^* [\rho^V(g) \otimes \rho^V(g)]$$

what remains is the two dimensional 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. 14.22 and 14.25 to show that the F matrix is given by

$$F_{VVf}^{VVd} = \frac{1}{2} \begin{pmatrix} 1 & 1 & \sqrt{2} \\ 1 & 1 & -\sqrt{2} \\ \sqrt{2} & -\sqrt{2} & 0 \end{pmatrix}$$

¹⁷ 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.

Temperly-Lieb Algebra and Jones-Kauffman Anyons

15

To give a definite example of the diagrammatic algebra related, 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, but we want to consider the case where all diagrams are planar — i.e., there are no over- and under-crossings. The only rule then is that the a loop is given a value d as shown in Fig. 15.1. As compared to the diagrammatic algebra we have constructed over the last few chapters (roughly starting in chapter 8, and continuing through chapter 11), 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 15.1 as a starting point. The algebra of loops that we will construct is known as the Temperly-Lieb algebra. While we will not consider over- and under-crossings yet, these be considered in chapter *** below and, taken together, the theories we construct here will be well defined anyon theories, sometimes known as Temperly-Lieb-Jones-Kauffman anyons, or just Kauffman anyons for short.

Let us start by thinking a bit about what kind of particle types we already have in our theory. Certainly we have the simple string¹ 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.

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. 15.2. This tells us immediately that

$$1 = \bar{1}.$$

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 as the quantum dimension 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. 15.3. This is a good idea, but it isn’t yet quite right. If 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

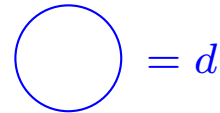


Fig. 15.1 The loop rule for the Kauffman bracket invariant and the Temperly-Lieb algebra.

¹It 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.

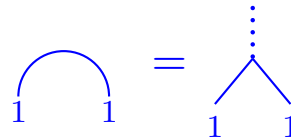


Fig. 15.2 Fusing two 1-particles to the vacuum

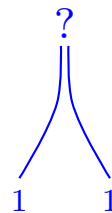


Fig. 15.3 Attempting to Fuse two 1-particles to something different from the vacuum

into another particle type (without fusing with some other particle or splitting). In Fig. 15.3 it looks like the two strings brought together could just fuse together to form the vacuum as in Fig. 15.2, and this would then turn the collection of two strings into the vacuum. To prevent such transmutation, we will work with operators known as projectors.

15.1 Jones-Wenzl Projectors



Fig. 15.4 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. 10.2). 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. 15.4). The Jones-Wenzl projector acting on the trivial case to consider is we will construct projectors out of two incoming 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. 15.3. 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 will construct the 2-string Jones-Wenzl projector P_2 out of two incoming 1-particles² (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 vacuum³ as shown in Fig. 15.5.

$$\bar{P}_2 = \frac{1}{d} \begin{array}{c} \cup \\ \cap \end{array} = \boxed{\bar{P}_2}$$

Fig. 15.5 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 bottom of the second operator. As shown in Fig. 15.6, 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.

²The Jones-Wenzl projector, if one defines one, for a single string is the trivial operator. I.e., one string comes in and the same string comes out unchanged.

³The estute reader will notice that a particle “turning around” as in Fig. 15.2 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 15.3 below.

$$[\bar{P}_2]^2 = \frac{1}{\bar{d}} \frac{1}{\bar{d}} \text{[loop]} = \frac{1}{d^2} \text{[two loops]} = \frac{1}{d} \text{[cup and cap]} = \bar{P}_2$$

Fig. 15.6 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. 15.7. 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.

$$P_2 = \text{[two parallel vertical lines]} - \frac{1}{d} \text{[cup and cap]} = \text{[labeled box } P_2 \text{]}$$

Fig. 15.7 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. 15.5 and 15.7. 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. 15.10 or the left of Fig. 15.8.

It is useful to calculate the quantum dimension of the 2-string⁴. This is shown in Fig. 15.8.

⁴The quantum dimension of the 2-string is called Δ_2 in many references (and similarly d_n is called Δ_n). We will stick with d to fit with the notation in the rest of this book.

$$d_2 = \text{[circle with label 2]} = \text{[circle with } P_2 \text{ box]} = \text{[double circle]} - \frac{1}{d} \text{[C-shape]} = d^2 - 1$$

Fig. 15.8 Evaluating the quantum dimension of the 2-string.

For $d = \pm 1$:

$$\begin{array}{|c|} \hline \\ \hline \end{array} \begin{array}{|c|} \hline \\ \hline \end{array} = \frac{1}{d} \begin{array}{c} \cup \\ \cap \end{array}$$

Fig. 15.9 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.

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. 15.9. In this case, notice that the projector $P_2 = 0$ since the two terms in the projector in Fig. 15.7 are equal with opposite signs. Correspondingly note that the quantum dimension of the putative 2-string is $d_2 = 0$ as shown in Fig. 15.8, meaning that no such particle exists.

The relevant theories here are bosons or fermions for $d = 1$ and semions for $d = -1$ (See Exercise 2.2). All of these theories are abelian, so it is not surprising that two particles which fuse to the identity cannot fuse together in a different way (recall that in abelian theories, the outcome of a fusion is always unique). The only possible outcome of fusion of two 1-strings is the vacuum as shown in Fig. 15.2. Thus the entire fusion rules of these theories are

$$1 \times 1 = 0$$

where again 0 is the identity or vacuum.

It is worth noting that the case of $d = -1$ falls into the category of theories we discussed in section 11.3 since the value of the loop is negative. As emphasized there, while we are free to consider diagrammatic algebras with $d = -1$ loops, this would correspond to a system with a non-positive-definite inner product (See Fig. 10.18), which is forbidden in quantum mechanics. Nonetheless, this can be made into a perfectly well behaved quantum theory by interpreting the string as a particle with negative Frobenius-Schur indicator $\kappa = -1$. Thus, if we want our diagrammatic algebra to represent a quantum theory, we must follow rule (0.a) and (0.b) from section 11.3 — that is, before evaluating a diagram we count the number of caps and call this number n ; then after evaluating the diagram we multiply the result by $(-1)^n$. The resulting, now well-behaved, diagrammatic algebra (including the additional rules (0.a) and (0.b)) corresponds to the semion, or $SU(2)_1$ theory.

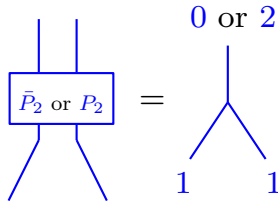


Fig. 15.10 Two possible fusions of two 1-strings, drawn in two different notations. A single line labeled 2 is interpreted as two 1-strings traveling together with a P_2 operator inserted. The label 0 means the two strands fuse to the vacuum as in Fig. 15.2.

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-strings can fuse to either 0 or 2 as shown in Fig. 15.10. 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. 15.11.

$$P_3 = \alpha \left| \begin{array}{c} | \\ | \\ | \end{array} \right| + \beta \left| \begin{array}{c} \cup \\ | \\ \cap \end{array} \right| + \gamma \left| \begin{array}{c} | \\ \cup \\ \cap \end{array} \right| + \delta \left| \begin{array}{c} \cup \\ \cup \\ | \end{array} \right| + \epsilon \left| \begin{array}{c} \cup \\ \cap \\ \cap \end{array} \right|$$

Fig. 15.11 The form of the most general three legged operator we can construct.

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. 15.12.

$$\begin{array}{c} 3 \\ | \\ \diagup \quad \diagdown \\ 1 \quad 0 \end{array} = \begin{array}{c} | \\ | \\ | \\ \boxed{P_3} \\ | \\ \boxed{\bar{P}_2} \\ | \\ | \end{array} = 0$$

Fig. 15.12 Insisting that 0×1 does not give 3

This and analogous constraints allow us to insist on the conditions shown in Fig. 15.13.

$$\boxed{P_3} = \boxed{P_3} = \boxed{P_3} = \boxed{P_3} = 0$$

Fig. 15.13 Four conditions that come from the fusion condition shown in Fig. 15.12.

However, we should allow fusions of the form $1 \times 2 = 3$ as shown in Fig. 15.14. Enforcing the condition in Fig. 15.13, along with $P_3^2 = P_3$ gives the form of P_3 shown in Fig. 15.11 with the results that (see Exercise 15.1)

$$\begin{aligned} \alpha &= 1 \\ \beta = \gamma &= -\frac{d}{d^2 - 1} \\ \delta = \epsilon &= \frac{1}{d^2 - 1} \end{aligned}$$

$$\begin{array}{c} 3 \\ | \\ \diagup \quad \diagdown \\ 1 \quad 2 \end{array} = \begin{array}{c} | \\ | \\ | \\ \boxed{P_3} \\ | \\ \boxed{P_2} \\ | \\ | \end{array}$$

Fig. 15.14 We allow $1 \times 2 = 3$

We can do a short calculation in the spirit of Fig. 15.8 to obtain the quantum dimension of the 3-string⁴, giving the result (See exercise 15.1) shown in Fig. 15.15.

$$d_3 = \bigcirc_3 = d(d^2 - 2)$$

Fig. 15.15 Evaluating the quantum dimension of the 3-string.

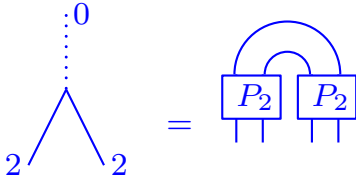


Fig. 15.16 $2 \times 2 = 0$.

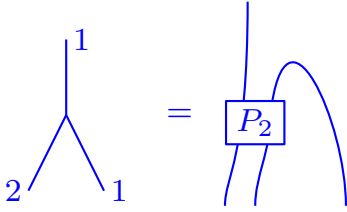


Fig. 15.17 $2 \times 1 = 1$. We recognize this as the fusion $1 \times 1 = 2$ from Fig. 15.10 just turned on its side.

For $d = \pm\sqrt{2}$:

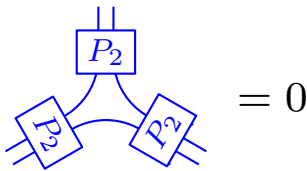


Fig. 15.18 For $d = \pm\sqrt{2}$ we have 2×2 not fusing to 2.

Ising Anyons

Consider the case where $d = \pm\sqrt{2}$. In this case, $d_3 = 0$ meaning that there is no 3-string particle. Equivalently it is possible to show that P_3 vanishes when evaluated in any diagram (See exercise 15.1). It is similarly possible to show that $P_4 = 0$ and so forth. Thus, in this theory 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. 15.16 and $2 \times 1 = 1$ as shown in Fig. 15.17. (Note that showing $2 \notin 2 \times 2$ requires another explicit calculation, not shown here! See exercise 15.1)

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 13.3) where $1 = \sigma$ and $2 = \psi$ and 0 is the vacuum I

Recall in our discussion of Ising anyons in section 8.2.2 we found that the quantum dimension of the σ particle (here the "1"-string) is $\sqrt{2}$ by studying the number of fusion channels for multiple σ anyons. As we found in section 12, this value of quantum dimension does indeed match the quantum dimension given by the value of a loop of 1-string.

For the case of $d = +\sqrt{2}$ the model here is precisely what is known as the Ising model, and it is very closely related to the Ising conformal field theory which we will encounter in section ***.

For the case of $d = -\sqrt{2}$ again we must use the approach of section 11.3. Here the 1-string, $d_1 = d = -\sqrt{2}$ represents a particle with negative Frobenius-Schur indicator $\kappa_1 = -1$, whereas the 2-string with $d_2 = d^2 - 1 = 1$ has positive Frobenius-Schur indicator $\kappa_2 = +1$. We can check that the fusion rules respect $\kappa_a \kappa_b = \kappa_c$ whenever $a \times b = c + \dots$ (where $\kappa_0 = +1$). The resulting diagrammatic algebra corresponds to the $SU(2)_2$ Chern-Simons theory.

15.2 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 15.2). 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. 15.19 (See exercise 15.2)

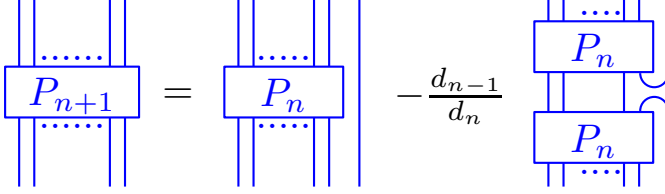


Fig. 15.19 Recursion Relation For Jones-Wenzl Projectors

Note in particular that if P_n vanishes, we can conclude that P_m vanishes for all $m > n$ as well.

We define the quantum dimension 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. 15.20.

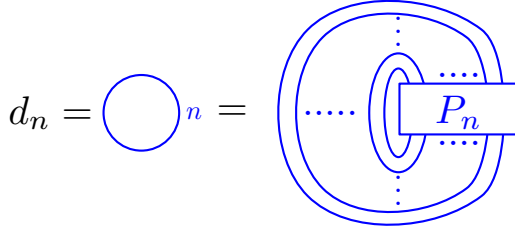


Fig. 15.20 Evaluating the quantum dimension of the n -string particle. We connect the n strings coming from the top of the projector P_n to those coming from the bottom. Often this quantity is notated as Δ_n .

Using the recursion shown in Fig. 15.19 and the definition of d_n in 15.20 we obtain the recursion relation

$$d_{n+1} = dd_n - d_{n-1} \quad (15.1)$$

where we define $d_{-1} \equiv 0$ and $d_0 = 1$ and hence $d_1 = d$. This recursion has the general solution

$$d_n = U_n(d/2) \quad (15.2)$$

where U_n is the n^{th} Chebyshev polynomial of the second kind. These are defined by (See exercise 15.2)

$$U_n(\cos \theta) \sin \theta = \sin[(n+1)\theta] \quad (15.3)$$

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 15.2)

$$d = 2 \cos \left(\frac{k\pi}{n+1} \right) \quad (15.4)$$

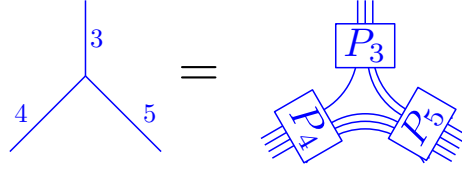


Fig. 15.21 The general vertex in the Temperley-Lieb algebra. Here the vertex is shown for 4 and 5 fusing to 3.

for some $k \in 1, \dots, n$. 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”).

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. 15.21. Consider a vertex between particle types (a, b, c) as in with $a, b, c, \geq 0$ as in Fig. 15.22. The number of strings going between the projectors (as in the right of Fig. 15.21) is given by

$$m = (a + b - c)/2 = \text{strings between } a \text{ and } b \quad (15.5)$$

$$n = (a + c - b)/2 = \text{strings between } a \text{ and } c \quad (15.6)$$

$$p = (b + c - a)/2 = \text{strings between } b \text{ and } c \quad (15.7)$$

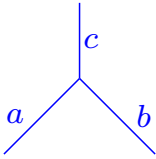


Fig. 15.22 A general vertex between particle types (a, b, c) with $a, b, c \geq 0$

these quantities must be non-negative, and we must have all of these quantities integer, which is assured if

$$(a + b + c) \text{ is even} \quad (15.8)$$

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

$$P_{(a+b+c)/2} \text{ is nonzero} \quad (15.9)$$

⁵This is entirely equivalent to the general $SU(2)_k$ Chern-Simons fusion rules where particles j take integer and half-integer values and

$$j_1 \times j_2 = |j_1 - j_2|, |j_1 - j_2| + 1, \dots, \min(j_1 + j_2, k - j_1 - j_2) \quad (15.10)$$

where we have made the identification that a in the Temperley-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. 15.10 is always $j_1 + j_2$), these fusion rules match the angular momentum addition rules of regular $SU(2)$.

This final condition is nontrivial and we will not prove it in all generality here (See for example, Kauffman and Lins [1994], for a proof). However, Fig. 15.18 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

$$a \times b = |a - b|, |a - b| + 2, \dots, \min(a + b, 2k - a - b)$$

where k is the largest integer such that P_k is non-zero.⁵

With this definition of a vertex we can evaluate diagrams. A particularly useful one is the version of the theta diagram shown in Fig. 15.23. The value of this diagram can be derived generally and is given by

$$\Delta(a, b, c) = (d_{(a+b+c)/2})! \frac{(d_{n-1})! (d_{m-1})! (d_{p-1})!}{(d_{a-1})! (d_{b-1})! (d_{c-1})!} \quad (15.11)$$

where we have defined

$$(d_n)! \equiv d_n d_{n-1} d_{n-2} \dots d_2 d_1$$

with $d_1 = d$ and $d_0 = d_{-1} = 1$. From Eq. 15.11 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. 15.9.

While the most general derivation of Eq. 15.11 is somewhat complicated (See Kauffman and Lins [1994]), it is easy enough to confirm it is correct for a few examples (See exercise 15.3)

The value, Eq. 15.11, of the Theta diagrams do not match what we would have expected given the rules in chapter 11. Comparing to Fig. 11.9 we would have expected the Theta diagram in Fig. 15.23 to have a value $\sqrt{d_a d_b d_c}$ which in general it does not.

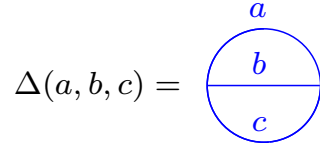


Fig. 15.23 The Theta diagram in the Temperley-Lieb-Jones-Kauffman theory.

15.3 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 established in prior chapters. In section ?? we just found that the value of the Theta diagram does not match the expectation from chapter 11. 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 11.

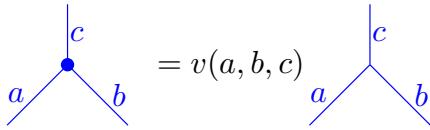


Fig. 15.24 A renormalized vertex between particle types (a, b, c) with $a, b, c \geq 0$ marked with a blue dot on the left is defined in terms of the original vertex on the right. We assume here that the vertex on the right, defined analogous to Fig. 15.21 is nonzero.

Let us define a new vertex which is a constant multiple of the old vertex as shown in Fig. 15.24. We define the rescaling factor as

$$v(a, b, c) = \sqrt{\frac{\sqrt{d_a d_b d_c}}{\Delta(a, b, c)}}$$

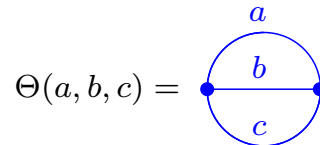


Fig. 15.25 The Theta diagram with renormalized vertices.

such that the value of the Theta diagram in Fig. 15.25 is now $\Theta(a, b, c) = \sqrt{d_a d_b d_c}$ as we expect from Fig. 11.9. It turns out that this simple modification is sufficient to make the theory fit into the framework developed in chapter 11.

15.4 F-matrices

We can now determine the F -matrices directly from the graphical algebra. As a simple example, consider the F -matrices $F_{11\beta}^{11\alpha}$ (which we abbreviate as F_{β}^{α}) as shown in Fig. 15.26. Note that for this equation we use renormalized vertices as defined in Eq. 15.24 and notated by dots on the vertices.

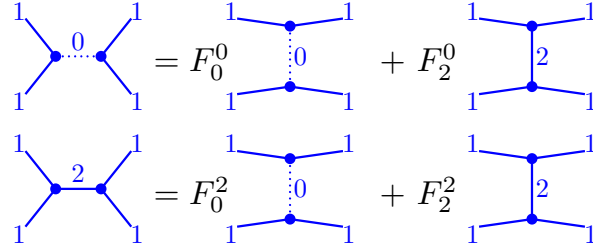


Fig. 15.26 The F -matrix in the Temperley-Lieb-Jones-Kauffman theory is unitary when use renormalized vertices, indicated by dots. Here we have abbreviated $F_{11\beta}^{11\alpha}$ as F_{β}^{α} for brevity.

This F -matrix equation is that of Fig. 11.1 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. 15.27. Note that in Fig. 15.27, 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.

$$\begin{aligned} \bigcup \bigcap &= F_0^0 \bigcup \bigcap + \frac{d}{\sqrt{d_2}} F_2^0 \left[\bigcup \bigcap - \frac{1}{d} \bigcup \bigcap \right] \\ \frac{d}{\sqrt{d_2}} \left[\bigcup \bigcap - \frac{1}{d} \bigcup \bigcap \right] &= F_2^0 \bigcup \bigcap + \frac{d}{\sqrt{d_2}} F_2^2 \left[\bigcup \bigcap - \frac{1}{d} \bigcup \bigcap \right] \end{aligned}$$

Fig. 15.27 Explicitly writing out the F -matrix equations of Fig. 15.26. 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 equa-

tions in Fig. 15.27. 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. 15.8) we obtain the full form of the F -matrix (and returning the 11 superscripts and subscripts which we have suppressed)

$$[F_{11}^{11}] = \begin{pmatrix} \frac{1}{d} & \frac{\sqrt{d^2-1}}{d} \\ \frac{\sqrt{d^2-1}}{d} & -\frac{1}{d} \end{pmatrix} \quad (15.12)$$

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. 13.25.

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 unrenormalized vertices.

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 15.1 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. 15.5 and 15.7.

(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 15.11.

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. 15.12 and 15.13.

Calculate the coefficients $\alpha, \beta, \gamma, \delta$ in P_3 . Calculate the quantum dimension d_3 shown in Fig. 15.15.

(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. 15.18.

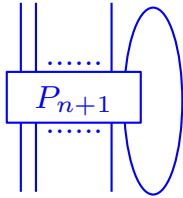


Fig. 15.28 This figure, with n strands going in the bottom, and n strands coming out the top, must be proportional to P_n .

Exercise 15.2 More General Jones-Wenzl Projectors

(a) A Jones-Wenzl projector for n strands is defined both by $P_n^2 = P_n$ as well as by being orthogonal to P_0 analogous to Fig. 15.13. Assuming these properties are satisfied for P_n show that they are satisfied for P_{n+1} given by Fig. 15.19. Hint: Use the fact that connecting up a single string from P_{n+1} from top to bottom as in Fig. 15.28 must give something proportional to P_n (Why?).

(b) Using Fig. 15.19 derive Eq. 15.1. Show that the solution to this equation is given by Eqs. 15.2 and 15.3. Confirm the condition for d_n to vanish given in Eq. 15.4.

Exercise 15.3 Theta Diagram

(a) Show $\Delta(a+1, a, 1) = d_{a+1}$. Hint: Use Fig. 15.28.

(b) More generally show $\Delta(a+k, a, k) = d_{a+1}$. Hint: Generalize Fig. 15.28 to the case where k strands are connected in a loop from the top to the bottom.

Exercise 15.4 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.

State Sum TQFTs

16

Having learned about planar diagrammatic algebras we are now in a position to explicitly construct a real 3D TQFT¹. 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 ***.

¹It is a bit surprising that one only needs a planar algebra to make a 3D TQFT!

16.1 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 dimensions².

16.1.1 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. 16.1.

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

²It 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 simplices.

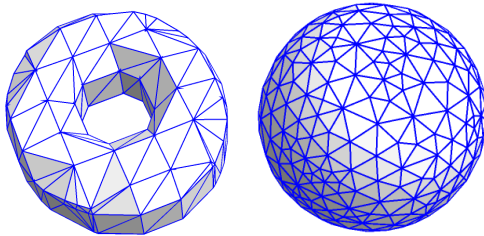


Fig. 16.1 Some triangulations of 2-manifolds

³I encourage you to play with these two moves and see how you can restructure triangulations by a series of Pachner moves.

⁴It is interesting to note that a Pachner move 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.

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. 16.2 and 16.3.

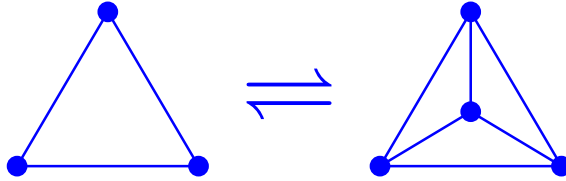


Fig. 16.2 The 1-3 Pachner move in two dimensions corresponds to adding or removing a point vertex from the triangulation. This turns one triangle into three or vice-versa.

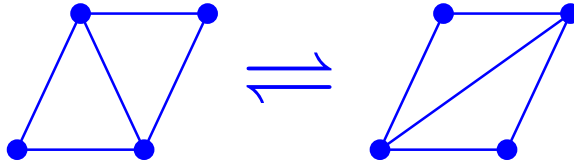


Fig. 16.3 The 2-2 Pachner move in two dimensions corresponds to replacing two adjacent triangles with two complementary triangles. This turns two triangles into two different triangles.

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.

16.1.2 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 manifold⁵ 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 moves⁶, which are shown in Figs. 16.4 and 16.5. 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.

⁵For now let us focus on closed manifolds. We briefly discuss manifolds with boundary in section 16.2.2.

⁶Analogous 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.

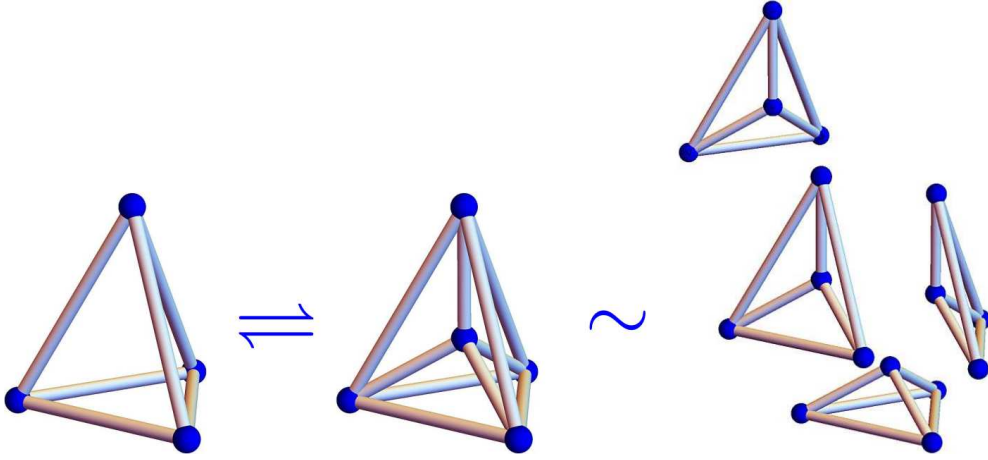


Fig. 16.4 The 1-4 Pachner move in three dimensions corresponds to adding or removing a point vertex to the tetrahedron decomposition. This turns a single tetrahedron into four or vice versa. On the far right we show the four tetrahedron separated for clarity.

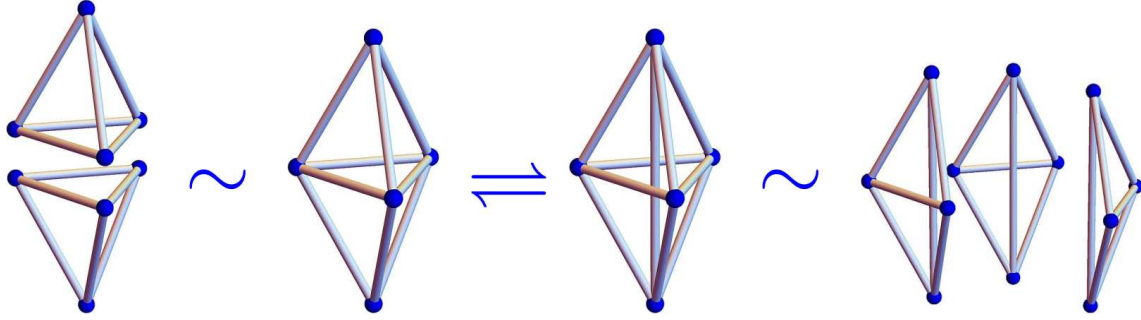


Fig. 16.5 The 2-3 Pachner move in three dimensions corresponds to re-splitting a double tetrahedron (left) into three tetrahedron (right). This turns a single tetrahedron into four or vice versa. On the far left we show the two tetrahedra separated for clarity; and on the far right we have the three tetrahedra separated for clarity.

16.2 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 8-15.

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 algebra⁷. We then consider the following sum

$$Z_{TV}(\mathcal{M}) = \mathcal{D}^{-2N_v} \sum_{\text{all edge labelings}} W(\text{labeling}) \quad (16.1)$$

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 known as the total quantum dimension (the sum is over all particle types). In Eq. 16.1, W is a weight assigned to each labeling of all the edges⁸. We consider the following definition of a weight assigned to a given labeling of edges

$$W(\text{labeling}) = \frac{\prod_{\text{tetrahedra}} \tilde{G}(\text{tetrahedron}) \prod_{\text{edges}} d_{\text{edge}}}{\prod_{\text{triangles}} \tilde{\Theta}(\text{triangle})} \quad (16.2)$$

Thus each tetrahedron is given a weight \tilde{G} , depending on its labeling, each edge is given a weight d which is its quantum dimension 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 studied⁹ in chapter 11 for example¹⁰. The functions \tilde{G} and $\tilde{\Theta}$ are given by

$$\tilde{\Theta} \left(\begin{array}{c} \text{triangle} \\ \text{with edges } c, b, a \end{array} \right) = \Theta(a, b, c) = \sqrt{d_a d_b d_c} \quad (16.3)$$

and

$$\tilde{G} \left(\begin{array}{c} \text{tetrahedron} \\ \text{with edges } d, c, a, b, e, f \end{array} \right) = G_{ecf}^{bad} = F_{ecf}^{bad} d_f \sqrt{\frac{d_b d_c}{d_f}} \sqrt{\frac{d_a d_e}{d_f}} \quad (16.4)$$

Note that the tetrahedron shown here is different from the one shown in Fig. 11.10 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. 11.10 the lines f, e, \bar{c} form a loop whereas f, \bar{e}, \bar{a} meet at a point. In the diagram in Eq. 16.4 on the other hand e, f, \bar{c} meet at a point where f, \bar{e}, \bar{a} form a loop. In Eq. 16.4 the three edges around any face must fuse together to the vacuum. I.e., we have the four conditions

$$N_{bad} > 0 \quad N_{c\bar{d}e} > 0 \quad N_{f\bar{e}\bar{a}} > 0 \quad N_{\bar{c}\bar{f}\bar{b}} > 0$$

or else \tilde{G} will vanish. Note that, like G , the value of \tilde{G} is unchanged under any rotation of the tetrahedron.

16.2.1 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 ***). Roughly, however, it is easy to see how it is going to work.

⁸In 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.

⁹Many works, including the original works by Turaev and Viro [1992], use the diagrammatic algebra based on Temperley-Lieb which we discussed in chapter 15. However, in those works, they have used the nonunitary version of the diagrammatic algebra without the vertex renormalization which we introduce in section 15.3. In such an approach $\Theta(a, b, c)$ is replaced by $\Delta(a, b, c)$, for example (See Eq. 15.11). 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!

¹⁰In chapter 11 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 (and pivotal) tensor category as we discuss in chapter 10. Full isotopy invariance is not required. This is discussed in depth by Barrett and Westbury [1996].

Let us first examine the 2-3 Pachner move shown in Fig. 16.5. 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. 16.4. After this cancellation of these factors, 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. 11.2 (or Eq. 9.3)! Thus any diagrammatic algebra which satisfies the pentagon equation will result in a Turaev-Viro invariant (Eq. 16.1) that is invariant under the 1-4 Pachner move!

The case of the 1-4 Pachner move is only a bit harder and we will sketch the calculation here. The large tetrahedron on the left of Fig. 16.4 (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) when 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. 11.4 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. 16.1 being that we have removed one vertex from the lattice in this procedure.

16.2.2 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 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. 16.1 where the sum is only over the edges in the bulk, leaving fixed (un-summed) the quantum numbers on the edges that live entirely on the boundary (i.e., both vertices on the boundary). Thus for manifolds with boundary we more generally write

$$Z_{TV}(\mathcal{M}; a_1, \dots, a_N) = \mathcal{D}^{-2N_v - n_v} W'(a_1, \dots, a_N) \sum_{\text{bulk labelings}} W(\text{bulk labels})$$

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. 16.2 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, \dots, a_N are the quantum numbers of the edges on the boundary, and these are not included in the sum over bulk labels. An additional weight is included which is a function of these boundary edge labels

$$W'(a_1, \dots, a_N) = \sqrt{\frac{\prod_{\text{boundary edges}} d_{\text{edge}}}{\prod_{\text{boundary triangles}} \tilde{\Theta}(\text{triangle})}}$$

The partition function $Z_{TV}(\mathcal{M}; a_1, \dots, a_N)$ is now a function the edge variables and is interpreted as a wavefunction $|Z(\mathcal{M})\rangle$ that lives on the boundary $\Sigma = \partial M$.

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, \dots, a_N of the edges on the boundaries. In other words, we have

$$\begin{aligned} Z_{TV}(\mathcal{M} \cup_{\Sigma} \mathcal{M}') &= \langle Z_{TV}(\mathcal{M}') | Z_{TV}(\mathcal{M}) \rangle \\ &= \sum_{a_1, \dots, a_N} [Z_{TV}(\mathcal{M}'; \bar{a}_1, \dots, \bar{a}_N)]^* Z_{TV}(\mathcal{M}; a_1, \dots, a_N) \quad (16.5) \end{aligned}$$

$$= \sum_{j_1, \dots, j_N} Z_{TV}(\mathcal{M}'; a_1, \dots, a_N) Z_{TV}(\mathcal{M}; a_1, \dots, a_N) \quad (16.6)$$

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 ***). The final result is easily seen to be the correct expression for the Turaev-Viro invariant for the full manifold $\mathcal{M} \cup \mathcal{M}'$.

As in section 7.2 one can generalize the idea of a TQFT to include particle world lines (labeled links) as 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)

16.3 Connections to Quantum Gravity Revisited

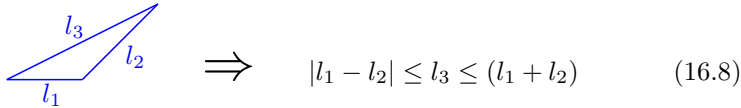
The Turaev-Viro invariant is a natural descendant of one of the very earliest approaches to quantum gravity pioneered by Penrose [1971] and Ponzano and Regge [1968]. Indeed much of the continued interest in Turaev-Viro and similar state-sum invariants is due to this relationship.

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 4D simplices (sometimes known as pentachora) in four dimensions¹¹. The curvature of the space-time manifold (the metric) is then determined by the lengths assigned to the edges¹².

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

$$Z = \int \mathcal{D}g \, e^{iS_{Einstein}[g]/\hbar} \quad (16.7)$$

We can imagine performing such a 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 inequality¹³



$$|l_1 - l_2| \leq l_3 \leq (l_1 + l_2) \quad (16.8)$$

The key observation is that the triangle inequality is precisely the same as the required inequality for regular angular momentum addition

$$j_1 \otimes j_2 = |j_1 - j_2|, |j_1 - j_2| + 1, \dots, |j_1 + j_2| \quad (16.9)$$

Thus it is natural to label each edge of with a quantum mechanical spin, and sum over all possible spins. Such an approach is known as a *spin network*. We thus imagine building a Turaev-Viro model (Eq. 16.1) with

¹¹It 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].

¹²All of general relativity can be reformulated in this discrete language. This is known as *Regge calculus*. See Regge [1961].

¹³With a bit of thought we realize that these inequalities must hold even with a curved spatial metric.

¹⁴Building a diagrammatic algebra based on a Lie group ($SU(2)$ in this case) is mentioned in section 14.2.2 above.

¹⁵Although the idea of spin networks as a toy model for quantum gravity goes back to Penrose [1971], and pursued further by Ponzano and Regge [1968], it was only much later that Hasslacher and Perry [1981] showed the precise equivalence of the model to gravity.

¹⁶As 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.

¹⁷Robbert 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 guttural sound that only exists in Dutch. However, those from the south of the Netherlands don’t use the guttural “g” and instead pronounce 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.

¹⁸For 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 define its value on a tetrahedron (see the comments in section 14.1.3).

¹⁹In chapter 14 we considered also the possibility of $d_a = -1$ but this is a gauge choice. We are always entitled to choose $+1$ instead at the cost of possibly losing isotopy invariance.

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. 16.9, and the F -matrices are given by the regular $6j$ symbols of angular momenta addition¹⁴. Such a model turns out to be very precisely¹⁵ the quantum gravity partition function Eq. 16.7 (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 15, see in particular margin note 5) implements this cutoff and yields a divergence-free result¹⁶.

16.4 Dijkgraaf-Witten Model

Another state sum model of some interest is the so-called Dijkgraaf-Witten model¹⁷ (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 14.1 and we use a 3-cocycle in place of the F -matrix¹⁸. These fusion rules require that multiplication of the group elements around every triangle must result in the identity as shown in Fig. 16.6. This is the analog of Eq. 16.3 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 ***.

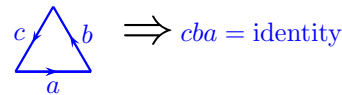


Fig. 16.6 Multiplying group elements around a triangle in Dijkgraaf-Witten theory results in the identity. This is known as the “flatness” condition

As mentioned in section 14.1 when we use group multiplication for fusion rules the quantum dimensions¹⁹ of all the particles are all $d_a = 1$. This means that in Eq. 16.2 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. 16.1 and 16.2 given by²⁰

$$Z_{DW}(\mathcal{M}) = |G|^{-N_v} \sum_{\text{labelings}} \prod_{\text{tetrahedra}} \tilde{G}(\text{tetrahedron}) \quad (16.10)$$

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. 16.6).

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

$$\tilde{G} \left(\begin{array}{c} i_1 \\ \diagup \quad \diagdown \\ i_2 \quad i_3 \\ \diagdown \quad \diagup \\ i_4 \end{array} \right) = \omega(g_{j_2, j_1}, g_{j_3, j_2}, g_{g_4, g_3})^{s(j_1, j_2, j_3, j_4)} \quad (16.11)$$

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 decomposing²¹. 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 ***).

²⁰With apologies for using G and \tilde{G} in the same equation to mean completely different things!

²¹To 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.

16.4.1 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, penta-chora 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. 16.1 — multiplying the group elements 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 -cocycle²² which we call $\omega_D(g_1, g_2, \dots, g_D)$ which is now a function of D arguments. Finally, one

²²I 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. 14.1 for the 3D case and Eq. 16.13 for the 2D case.

builds a partition function by summing over all possible labelings

$$Z_{DW}(\mathcal{M}_D) = |G|^{-N_v} \sum_{\text{labelings}} \prod_{D\text{-simplices}} \omega_D(g_{j_2, j_1}, \dots, g_{j_{D+1}, j_D})^{s(j_1, \dots, j_{D+1})} \quad (16.12)$$

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 $j_1 < j_2 < \dots < j_D$. Finally the exponent s is always ± 1 depending on whether the orientation of simplex described by the ordered set $[j_1, \dots, j_{D+1}]$ 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 condition²³

$$\omega_2(g, h)\omega_2(gh, k) = \omega_2(h, k)\omega_2(g, hk) \quad (16.13)$$

In the partition function, Eq. 16.12, 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. 16.7.

²³The 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 28.2.4.

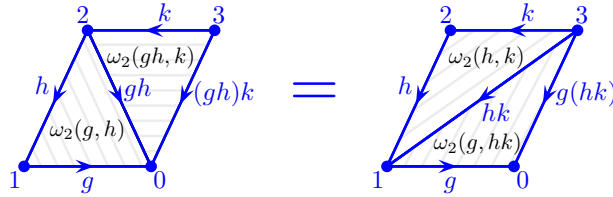


Fig. 16.7 Each triangle satisfy the flatness condition Eq. 16.6 meaning multiplying all three edges in order gives the identity. In the partition function each triangle gets a weight given by the corresponding cocycle ω_2 as written in black text. All of the triangles in the figure are oriented positively $s = +1$. The cocycle condition Eq. 16.13 guarantees that the product of the cocycles on the left equals the product of the cocycle on the right.

16.4.2 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. 16.10 (or more generally Eq. 16.12) is just unity so the partition function just counts the number of flat field configurations (See Fig. 16.6) and then divided by $|G|^{N_v}$. This partition function is exactly that of lattice gauge theory, as we will see in chapter *** below. 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.

The 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 Dijkstra-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!
- Recent book Turaev and Virelizier [2017]
- Discrete Gravity Regge and Williams [2000]
- Regge original work Regge [1961]
- Ooguri [1992]
Crane and Yetter [1993]
spin foam Lorente [2006].

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, and extensions to Lorentzian signature. The rather

Exercises

Exercise 16.1 Some 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, \dots, j_N . Assume tha the theory has reflection symmetry as in Eq. 11.13, show that

$$[Z_{TV}(\mathcal{M}; \bar{a}_1, \dots, \bar{a}_N)]^* = Z_{TV}(\mathcal{M}; a_1, \dots, a_N)$$

And as a result show that for a closed manifold $Z(\mathcal{M})$ is real.

Braiding and Twisting

17.1 Twists

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

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

In our diagrammatic notation, we have twist factors defined by Fig. 17.1. We note that in many cases, quantities of interest will depend only on the twist factor, i.e., the fractional part of the topological spin, $h_a \bmod 1$. It is often hard to pin down the value of the topological spin itself.

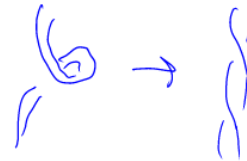


Fig. 17.2 Pulling a ribbon straight

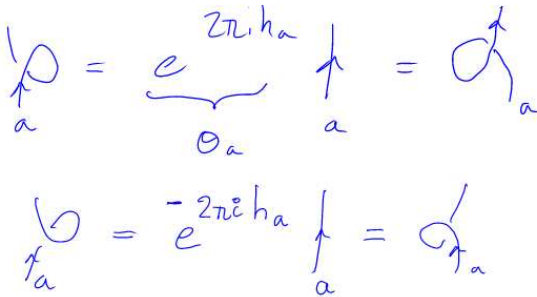


Fig. 17.1 Definition of Twist Factor

Recall we should treat particle world-lines as ribbons, so that a loop can be pulled straight as in Fig. 17.2 to represent a particle twisting around its own axis, as well as giving the phase of exchange for two identical particles (See also Fig. 2.7). Two cases are well known to us: if the spin h_a is an integer, then $e^{2\pi i h_a}$ is the identity, and this particle is a boson. If h_a is a half-odd-integer, then the phase is -1 and the particle is a fermion. The vacuum, or identity particle, should have zero scaling dimension, $h_I = 0$.

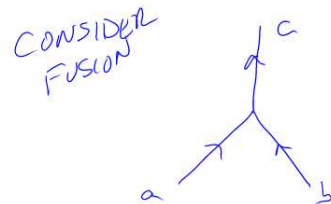


Fig. 17.3 Two particles fusing to a third. For this anyon system $a \times b = c + \dots$, and c is the particular fusion channel that has occurred in this diagram.

17.2 R-matrix

Consider the possibility of two particles fusing to a third as shown in Fig. 17.3. We have $a \times b = c + \dots$. I.e., c is among the possible fusion channels that can occur and we assume in the diagram that c

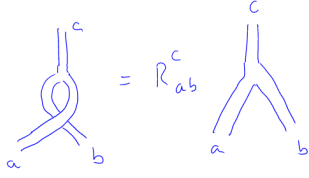


Fig. 17.4 Definition of R -matrix. Here we drop the arrows for convenience of notation and we draw the particle world lines as ribbons to show that no additional self-twists are incurred by the particles.

is the particular fusion channel that has occurred. Now let us consider braiding a and b around each other before fusing them as in Fig. 17.4. This diagram defines the so-called R -matrix. Here we have dropped the arrows and we show the particle world lines as ribbons to show that there are no additional self-twists. Note that braiding anything with the identity particle should be trivial, $R_{Ia}^a = R_{aI}^a = 1$.

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

To see the relationship between braiding and twisting, consider applying the R matrix twice to make a double twist as in Fig. 17.5. By pulling tight the double twist, the diagram can be reduced to twist factors previously defined, and this fixes R_{ab}^c up to a possible minus sign.

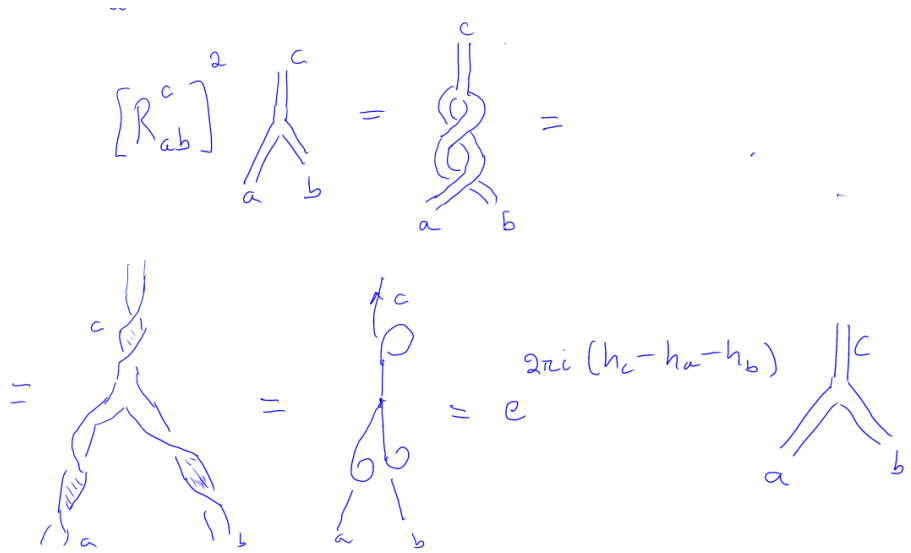


Fig. 17.5 Relation of R -matrix to twist factors. Note there is an error in the picture. It should read $R_{ba}^c R_{ab}^c$

We can generally write this relationship as

$$R_{ba}^c R_{ab}^c = e^{2\pi i(h_c - h_a - h_b)} = \theta_c / (\theta_a \theta_b) \quad (17.1)$$

Example: Fibonacci Anyons

In the Fibonacci theory, two τ particles can fuse to either τ or I . Applying the above relationship, Eq. 17.1, we have

$$[R_{\tau\tau}^{\tau}]^2 = e^{2\pi i(h_{\tau} - h_{\tau} - h_{\tau})} = e^{-2\pi i h_{\tau}} \quad (17.2)$$

$$[R_{\tau\tau}^I]^2 = e^{2\pi i(h_I - h_{\tau} - h_{\tau})} = e^{-4\pi i h_{\tau}} \quad (17.3)$$

Using the F and R matrices for a general anyon theory we can evaluate the unitary transform associated with any braid. Recall the two possible states of three τ particles fusing to τ as shown in Fig. 17.6.

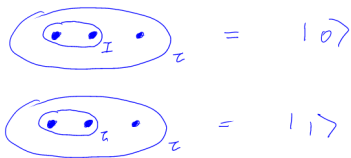


Fig. 17.6 The two states of three τ particles fusing to τ . Unmarked dots are τ particles.

Now consider braiding the two leftmost particles around each other.

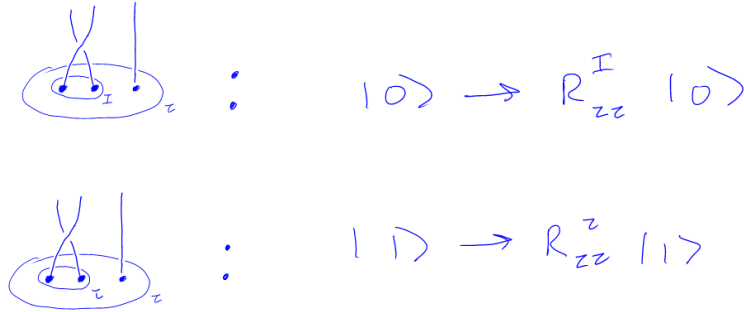


Fig. 17.7 Braiding the two left particles in this basis gives a phase dependent on the fusion channel of the two particles.

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

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

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

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

Now how can we evaluate the braid shown in Fig. 17.8? The trick here is to use the F -matrix to change the basis such that we know the fusion channel of the right two particles, and then once we know the fusion channel we can use the R -matrix. If we want, we can then use the F -matrix to transform back to the original basis. To see how this works, Recall that we can use the F matrix to write (See Eq. 9.1)

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

or in diagrams (see Fig. 17.9).



Fig. 17.9 The F -matrix relation in diagram form. See Eq. 9.1

On the right hand side of Fig. 17.9 (i.e., in the prime basis) we know the fusion channel of the rightmost two particles, so we can braid them around each other and use the R -matrix to accumulate the corresponding phase.

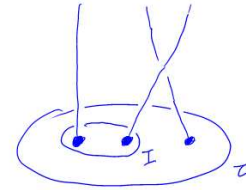


Fig. 17.8 How does one evaluate this braid? One applies F -first, then R as shown in the next two figures!

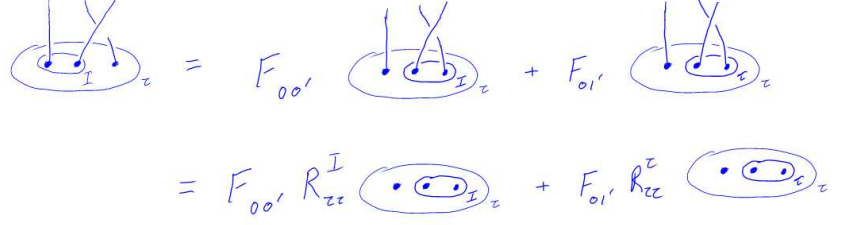


Fig. 17.10 To braid particles, switch basis using F until we know the fusion channel of the two particles we want to braid, and then we can apply the R -matrix.

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

$$\begin{aligned} R_{23}|0\rangle &= R_{23}(F_{00'}|0'\rangle + F_{01'}|1'\rangle) \\ &= F_{00'}R_{23}|0'\rangle + F_{01'}R_{23}|1'\rangle \\ &= F_{00'}R_{\tau\tau}^I|0'\rangle + F_{01'}R_{\tau\tau}^\tau|1'\rangle \end{aligned} \quad (17.4)$$

$$\begin{aligned} &= F_{00'}R_{\tau\tau}^I ([F^{-1}]_{0'0}|0\rangle + [F^{-1}]_{0'1}|1\rangle) \\ &\quad + F_{01'}R_{\tau\tau}^\tau ([F^{-1}]_{1'0}|0\rangle + [F^{-1}]_{1'1}|1\rangle) \\ &= (F_{00'}R_{\tau\tau}^I [F^{-1}]_{0'0} + F_{01'}R_{\tau\tau}^\tau [F^{-1}]_{1'0}) |0\rangle \\ &\quad + (F_{00'}R_{\tau\tau}^I [F^{-1}]_{0'1} + F_{01'}R_{\tau\tau}^\tau [F^{-1}]_{1'1}) |1\rangle \end{aligned} \quad (17.5)$$

¹For this particular case (using Eq. 9.2 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!)

Where between Eq. 17.4 and 17.5 we have used the inverse F transform to put the result back in the original $|0\rangle$ and $|1\rangle$ basis.¹

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

17.3 The Hexagon

As with the case of the F -matrix, there are strong consistency constraints on the R -matrices given a set of F -matrices (indeed, it is possible that for a given set of F -matrices that satisfy the pentagon, there may not even exist a set of consistent R -matrices!). The consistency equations are known as the hexagon equations and are shown diagrammatically in Fig. 17.11. In equations this can be expressed as

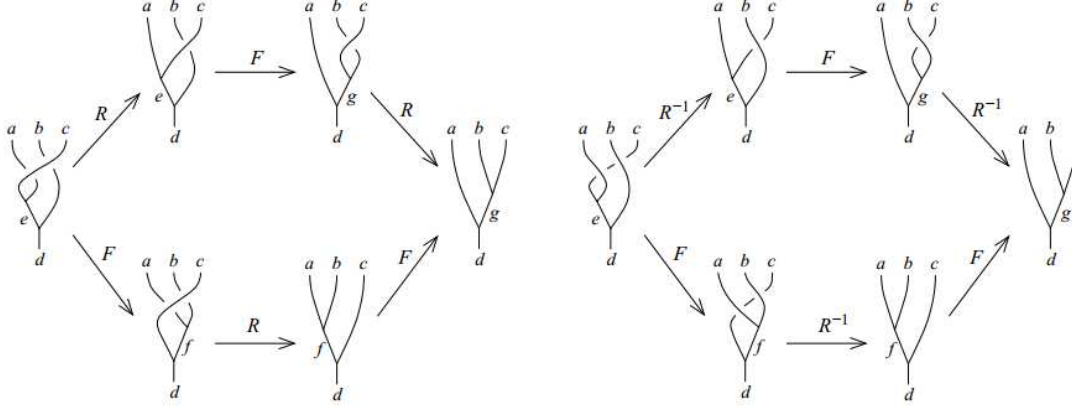


Fig. 17.11 The hexagon equations in graphical form. (nice picture stolen from Bonderson thesis)

$$\begin{aligned}
 R_e^{ca} [F_d^{acb}]_{eg} R_g^{cb} &= \sum_f [F_d^{cab}]_{ef} R_d^{cf} [F_d^{abc}]_{fg} \\
 [R_e^{ca}]^{-1} [F_d^{acb}]_{eg} [R_g^{cb}]^{-1} &= \sum_f [F_d^{cab}]_{ef} [R_d^{cf}]^{-1} [F_d^{abc}]_{fg}
 \end{aligned}$$

The top equation is the left diagram whereas the lower equation is the right diagram in Fig. 17.11. The left hand side of the equation corresponds to the upper path, whereas the right hand side of the equation corresponds to the lower path.

In simple theories such as the Fibonacci theory, knowing the F matrices, the Hexagon equation almost uniquely defines the R -matrices. In fact there are two consistent solutions to the Hexagon equations for the Fibonacci theory (See exercise 17.1).

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

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

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

17.4 Ocneanu Rigidity

Given a set of fusion rules, the pentagon and hexagon equation are very strong constraints on the possible F and R matrices that can result. (For example, as mentioned above, with Fibonacci fusion rules, there is

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

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

⁴It is often useful to impose one more condition, that the theory is “modular” which we will discuss below in section ***. Most well behaved theories are modular, although the presence of a fermion makes a theory non-modular — indicating how difficult it is to properly treat fermions! As far as we can tell from the known periodic table, all modular theories can be described in terms of some sort of Chern-Simons theory or closely related construction!

only one solution of the pentagon up to a gauge freedom and then only two solutions the hexagon). In fact, it is a general principle that the pentagon and hexagon for any set of fusion rules for a finite set of particles will have a finite set of solutions. In particular, once we have a set of solutions, in no sense is there a way that we can deform the values of F and R by a small amount and have another solution. This is a principle known as *rigidity* of the solutions, and it was first pointed out by Ocneanu². This principle makes it possible to contemplate putting together a sort of “periodic table” of possible anyon theories, starting with those having very few particle types. In fact, such periodic tables have been compiled up to about five or six different particle types^{3,4}. There is nothing in principle that prevents one from listing all the possible anyon theories even for more particle types although the search for all solutions becomes extremely difficult for greater numbers of particles.

17.5 Appendix: Gauge Transforms and R

As in section 9.1.3 one can gauge make general gauge transformations on the vertices of a theory. Given the transform

$$\begin{array}{c} a \quad b \\ \diagdown \quad \diagup \\ \mu \\ \diagup \\ c \end{array} \rightarrow u_{c;\mu}^{ab} \begin{array}{c} a \quad b \\ \diagdown \quad \diagup \\ \mu \\ \diagup \\ c \end{array}$$

the R matrix transforms as

$$R_c^{ab} \rightarrow \frac{u_c^{ba}}{u_c^{ab}} R_c^{ab}$$

Note that R_c^{aa} is gauge invariant, as is $R_c^{ab} R_c^{ba}$ in Eq. 17.1.

17.6 Appendix: Parastatistics Revisited

Given that we have now explained the general topological structure of theories in 2+1 dimensions we re-ask the question of whether we can have exotic statistics of particles in 3+1 dimensions. We discussed this issue previously in section 3.5.1. Recall that in order to construct a path integral we were looking for the topologically distinct classes of paths through configuration space — which we identified as the braid group in 2+1 but the symmetric group in 3+1. The idea of using a nonabelian representation of the symmetric group is known as parastatistics. We claimed that if we are to allow particle creation/annihilation, and we enforce locality, such types of statistics can be excluded.

Now that we have explored the topological structure of general theories in 2+1 dimension, we can try to see whether a similar structure

can be imposed upon a 3+1 dimensional theory. In 3+1 dimensions, recall that one cannot distinguish an over-crossing from an undercrossing (See section 3.3.2) and a double exchange is equal to the identity. Thus we have the gauge invariant statement $R_c^{ab} R_c^{ba} = 1$. In particular this implies $R_c^{aa} = \pm 1$. Note that we leave open the possibility that two particles a and b may fuse to multiple results c .

Further we can look at the 3+1 dimensional analogue of a particle twist as in Fig. 17.1. Since over and undercrossings are identical, we conclude that the twist factor is $\theta = \theta^{-1} = \pm 1$. Equivalent to this statement we consider a coordinate frame (x, y, z) which may rotate as a function of time (analogous to rotating a vector as a function of time in 2 spatial dimensions, which traces out a ribbon in 2+1 d). Since each rotation is an element of the group $SO(3)$ and we know that two full rotations in $SO(3)$ is topologically equivalent to the identity (but one rotation is not! See section 28.3.1), we conclude that the phase associated with a full rotation must be ± 1 (corresponding to integer and half integer spins).

Let us further consider $a \times b = c + \dots$ then we can consider a full rotation of c as being a full rotation of a and a full rotation of b combined with braiding a all the way around b . Since the full braiding of a around b is trivial in 3+1 dimension (i.e., since $R_c^{ab} R_c^{ba} = 1$) we may conclude that $\theta_a \theta_b = \theta_c$. This can also be seen from Eq. 17.1 along with $R_c^{ab} R_c^{ba} = 1$.

Thus, we have a \mathbb{Z}_2 grading of our fusion algebra, where particles are classified as having $\theta = \pm 1$ (obviously these are going to correspond to bosons or fermions).

Further reading

This is some reading

Exercises

Exercise 17.1 Fibonacci Hexagon In any TQFT or "braided" (including modular) tensor category (think of all of these as just anyon theories! don't worry about the fine distinctions), a braiding is defined by an R-matrix as shown in the figure 17.4. Once F matrices are defined for a TQFT, consistency

of the R -matrix is enforced by the so-called hexagon equations as shown in the figure diagrammatically by Fig. 17.11.

For the Fibonacci anyon theory, once the F matrix is fixed as in Eq. 9.2, 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.

Diagrammatic Algebra, the S -matrix and the Verlinde Relation

18

We have built up our anyon theories and now, using F and R matrices we can generally figure out how the degenerate Hilbert space $V(\Sigma)$ evolves (where by Σ we mean a surface with particle in it) as the particles move around in the manifold.

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

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

where here the world lines should be allowed to fuse and split, and as discussed in section?? one can always choose the manifold to be S^3 or some other simple reference manifold if one is willing represent other manifolds via Kirby calculus.

Note that while the diagrammatic calculus for the Kauffman case is often quite simple, there can be some nasty bookkeeping glitches for other anyon theories. For careful details of how all of the details, see Kitaev 2005 or Bonderson's thesis (** See also chapter ** to be added).

First, we should be careful about our normalization when we evaluate some knot or link of world lines¹. We choose our evaluation of a world line link to be of the form

¹We allow branching world lines which correspond to fusion or splitting

$$\langle \text{Link} \rangle \equiv \frac{Z(S^3 \text{ with embedded Link})}{Z(S^3)} \quad (18.1)$$

$$= Z(S^2 \times S^1 \text{ with embedded Link}) \quad (18.2)$$

where in the case of $S^2 \times S^1$ we require that the Link not go around the nontrivial handle of the S^1 . This normalization is chosen so that the evaluation of the empty link will give unity (as discussed in chapter 7).

$$\langle \emptyset \rangle = 1 \quad (18.3)$$

By using F 's and R 's we hope to reduce diagrams to a collection of non-linking labeled loops (labeled with their particle type), similar to what we did in evaluating the Kauffman invariant. We then need to know what value to give a particular loop.

18.1 Normalizations and Loops

² The reason we should assume that this is a positive quantity is because, cutting the loop half-way up, it can be thought of as an inner product between two identical states $\bigcirc = \langle \cup | \cup \rangle = d_a$ (most directly we can see that the two halves are identical when they are embedded in $S^2 \times S^1$ since the manifold is assembled by gluing two identical pieces as well.). While the value of this states is not normalized to unity in our convention, we will insist that we have postive-normed states.

³ In some cases it is convenient to define the value of a loop to be negative, as in the case of the semion Kauffman theory discussed above. However, by redefining some F -matrix elements, one can always work with the convention that d_a is positive, although this comes at the expense of having troublesome minus signs pop up in other places! These minus signs are known as Frobenius-Schur indicators and will be discussed in section ***.



Fig. 18.1 A loop of particle type a is given value $d_a > 0$. This will turn out to be the quantum dimension of the particle.

⁴It turns out that any fermion will make a theory non-modular! This is why fermions are a bit difficult to handle!

⁵See footnote 16 in chapter 7 for why S_{00} must be real.

Let us define $d_a > 0$ to be the value associate with the a loop of particle of type a as shown in Fig. 18.1^{2,3}. Note that we have $d_0 = 1$ for the identity particle (due to Eq. 18.3).

These quantities will turn out to be the quantum dimensions of the particles, but we have not shown this yet! We have not yet decided what value this loop should get. However, we can look back to 7.6 to note that we have

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

where S_{ab} is the unitary matrix known as the modular S -matrix. Recall that S should be unitary because it can be interpreted as a change of basis. (Theories where the S matrix comes out non-unitary are considered badly behaved, or “non-modular”. We will ignore this harder case for now⁴).

We can then think of the single loop d_a as particle a linking the vacuum, so we write

$$Z(S^3; a \text{ loop}) = S_{a0} = S_{0a}$$

and further we can write the normalizing factor $Z(S^3)$ as vacuum linking vacuum, so we have the value of a single loop as

$$d_a = S_{a0}/S_{00}$$

The fact that S is unitary gives us a useful identity

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

where the sum is over all particle types in the theory. We can then write⁵

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

where \mathcal{D} is known as the total quantum dimension and is given by

$$\mathcal{D}^2 = \sum_a d_a^2$$

Note that, as of this point we still have not shown that the d_a ’s, i.e., the values of the loops, are related to the quantum dimensions.

18.2 Quantum Dimensions

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

$$d_a d_b = \sum_c N_{ab}^c d_c. \quad (18.4)$$

Diagrammatically we have Fig. 18.2 This rule seems rather natural,



Fig. 18.2 The quantum dimensions satisfy the fusion algebra. See rigorous derivation in chapter appendix ***

that a and b can fuse together to form c in all possible ways (See also Exercise 8.1). However, to prove it is a bit more complicated than this argument, and is given in the appendix to this chapter.

Now, given Eq. 12.1, if we think of the fusion multiplicity for particle a , as a matrix N_a with indices b and c , and we think of d_c as a vector \vec{d} we can write

$$d_a \vec{d} = [N_a] \vec{d}$$

I.e, the vector \vec{d} is an eigenvector of N_a with eigenvalue d_a .

Note that the matrix N_a has only non-negative elements and \vec{d} has only positive elements. This allows us to apply the Perron-Frobenius theorem which says that for matrices with only non-negative elements⁶ there is a unique eigenvector with all positive entries, and it corresponds to the largest eigenvalue. Thus we conclude that d_a is actually the largest eigenvalue of the matrix N_a and it has eigenvector \vec{d} .

Recall that our previous definition of the quantum dimension d_a is that it is the largest eigenvalue of the fusion multiplicity matrix N_a . Thus we have rigorously shown that the value d_a of the loop in the graphical algebra is precisely the quantum dimension!

18.3 Verlinde Algebra

Using the locality principle (or no-transmutation) principle (See Fig. 8.7) we can show that a closed loop of type a around a world line of type x gives some constant which we call \tilde{S}_{ax} as shown in Fig. 18.3. by bending the top of x and forming a closed loop with the bottom of x , we construct linked rings on the left of this equation which we relate to the modular S -matrix, but on the right we form just a single x -loop.

$$S_{ax} = Z(S^3, a \text{ loop links } x \text{ loop}) = \tilde{S}_{ax} Z(S^3, x\text{-loop}) = \tilde{S}_{ax} S_{0x}$$

from which we conclude

$$\tilde{S}_{ax} = \frac{S_{ax}}{S_{0x}} \quad (18.5)$$

On the other hand, if we have two loops a and b around x , we can fuse the two loops to all possible loops c as shown in Fig. 18.4. This identity is entirely analogous to that of Fig. 18.2, and the rigorous derivation is given in the appendix. On the other hand, we could also evaluate the left hand side of Fig. 18.4 by applying the identity of Fig. 18.3 twice in a row, and similarly we can evaluate the right hand side of Fig. 18.4 by applying Fig. 18.3 once. Thus we obtain the identity

$$\tilde{S}_{ax} \tilde{S}_{bx} = \sum_c N_{ab}^c \tilde{S}_{cx}$$

This important result can be re-presented in two important ways. First, inverting this matrix equation gives

$$N_{ab}^c = \sum_x \tilde{S}_{ax} \tilde{S}_{bx} [\tilde{S}^{-1}]_{xc}$$

⁶Actually the simplest version of Perron-Frobenius requires all positive elements. Using the theorem for non-negative matrices allows there to be a second eigenvalue with same magnitude but opposite sign — this does not change the conclusion.

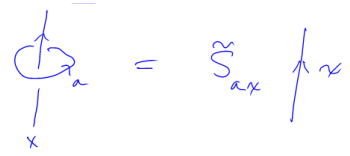


Fig. 18.3 The locality principle tells us that the value of a loop around a world line is some number which we call \tilde{S}_{ax}

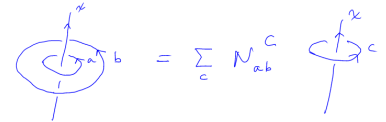


Fig. 18.4 Similar reasoning as in Fig. 18.2 allows us to write this diagrammatic relationship. See rigorous derivation in chapter appendix ***

Plugging in the value of \tilde{S} from Eq. 18.5, and using the fact that the modular S matrix is unitary, we obtain the famous Verlinde formula

$$N_{ab}^c = \sum_x \frac{S_{ax} S_{bx} S_{cx}^*}{S_{0x}}$$

which tells us that all the information about the fusion algebra is contained entirely within the modular S matrix!

A second way to present this important results is to write it in the form

$$[S^\dagger N_a S]_{xy} = \tilde{S}_{ay} \delta_{xy}$$

where N_a here is the matrix N_{ab}^c with indices b and c . Thus the result tells us that the modular S matrix is precisely the unitary diagonalizing matrix we were looking for in Eq. 8.11!

18.4 Return of Kirby Color

As mentioned in section 19.2.3, one can assemble a string called the “Kirby Color” (or Ω string) that is the sum of all strings weighted by the S -matrix.

$$|\Omega\rangle = \sum_a S_{0a} |a\rangle = \frac{1}{\mathcal{D}} \sum_a d_a |a\rangle \quad (18.6)$$

This string has some remarkable properties. Suppose we loop this string around a string x similar to that of Fig. 18.3. The result then looks like

$$\sum_a S_{0a} \tilde{S}_{ax} |x\rangle = \mathcal{D} \delta_{x0} |0\rangle$$

where we have used the fact that S is unitary, that $S_{0a} = S_{a0}$ is real, and that $S_{00} = 1/\mathcal{D}$. This is shown explicitly in fig. 18.5

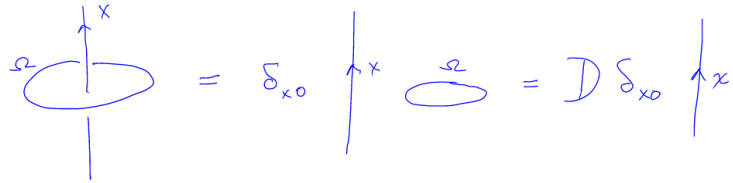


Fig. 18.5 The killing property. A loop of the Kirby color string projects to the vacuum going through it.

Thus, a loop of Kirby color string projects to zero (or vacuum) flux going through it! This is sometimes known as the “killing property”, as a loop of Ω string kills any non-trivial particle that tries to go through it. This principle is extremely useful in later attempts to construct topological models.

Further, the Kirby color string can be used, as mentioned in section 19.2.3 to build up a manifold invariant from anyon braiding rules. Indeed we can check this. The evaluation of the empty knot is defined to be $1 = Z(S^3)/Z(S^3)$. Starting with S^3 , surgery on a single loop takes S^3 to $S^2 \times S^1$. We thus expect that

$$\langle \text{single } \Omega\text{Loop} \rangle = \frac{Z(S^2 \times S^1)}{Z(S^3)} = \frac{1}{S_{00}} = \mathcal{D}$$

where we have used that $Z(S^2 \times S^1) = 1$ and $Z(S^3) = S_{00} = 1/\mathcal{D}$. Now, let us evaluate the Ω -loop using our diagrammatic rules as shown in Fig. 18.6.

$$\sum_a \frac{d_a}{D} \text{ (loop with twist } \theta \text{ and index } a \text{)} = \sum_a \frac{d_a^2}{D} = D$$

Fig. 18.6 The value of an Ω loop is \mathcal{D} .

Indeed, this gives \mathcal{D} in agreement with our surgery prediction. So this appears to be working! One should be a bit careful with this because one needs to properly account for twists in loops which we have not done here. See the more detailed discussion in section ***.

18.5 S and θ and that is all?

In building up an anyon theory, we now have compiled a large amount of data. Say there are M particle types, then we have F matrices, which have 6 indices⁷, each running from 1 to M , we have N matrices with three indices, we have R matrices with three indices, we have S matrices with two indices, and d 's and θ both with one index each. This seems like a huge amount of data needed to keep track of (and in some sense it is a huge amount of data). However, due to the idea of Rigidity (see section 17.4), it is believed that you need only specify the matrix S_{ab} and the values of the twists θ_a and you completely pin down the rest of the theory! This statement is not proven, but there are no counter-examples known.

⁷Here we assume no fusion multiplicity greater than 1. If we have such multiplicities, we would add additional indices to the F -matrices.

18.6 Appendix: Quantum Dimensions Satisfy the Fusion Algebra

We would like to show the identity shown in Fig. 18.2. We need a few useful pieces. First note that we can use an F -move on parallel lines to show the identity shown in Fig. 18.7.

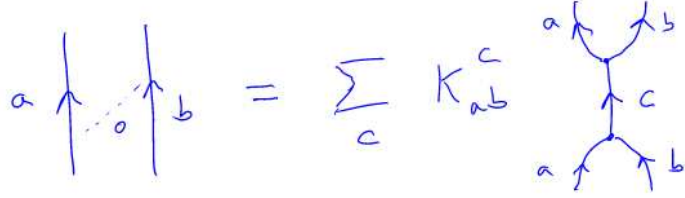


Fig. 18.7 An F-move. If a and b do not fuse to c , then the coefficient κ_{ab}^c must be zero. And if a and b do fuse to c then κ_{ab}^c is not zero. Note that the constant κ_{ab}^c shown here is typically notated as F_{abc}^{ab0} . This is quite similar to $[F_{abb}^a]_{0c}$ except that some lines pointing up have been turned down. This incurs certain normalization factors that one needs to keep track of. See section ***

Further we can use the locality principle (See Fig. 8.7) to give us Fig. 18.8

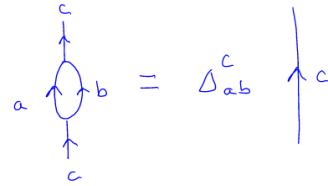


Fig. 18.8 Removal of a bubble gives a factor, which we call $\Delta_{ab}^c \neq 0$.

We can then use these two identities to directly fuse the loop of a with the loop of b incurring a factor of $\kappa_{ab}^c \Delta_{ab}^c$ as shown in Fig. 18.9

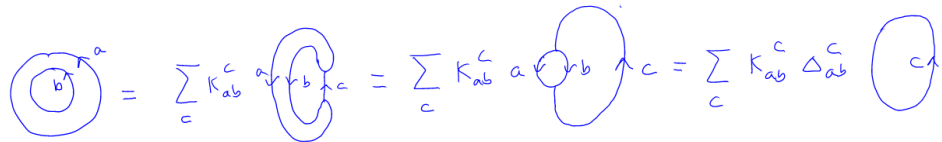


Fig. 18.9 We have applied first the result of Fig. 18.7 then Fig. 18.8. Note that if a and b cannot fuse to c then that term is zero in the sum.

However, we can also apply the same reasoning to split the loops into multiple bubbles as shown in Fig. 18.10.

$$\text{Diagram 1} = \sum_c (K_{ab}^c)^2 = \sum_c (K_{ab}^c \Delta_{ab}^c)^2 = \text{Diagram 2}$$

Fig. 18.10 Applying the result of Fig. 18.7 twice then Fig. 18.8 twice.

From these two results we can immediately conclude that $\kappa_{ab}^c \Delta_{ab}^c = 0$ or 1. Since both of these factors are nonzero when a and b can fuse to c , and are zero when they cannot, we can write $\kappa_{ab}^c \Delta_{ab}^c = N_{ab}^c$ (assuming no $N_{ab}^c > 1$)⁸. This then proves our Lemma.

Once it is established that the factor $\kappa_{ab}^c \Delta_{ab}^c = N_{ab}^c$ then this can be also used to directly prove the identity in Fig. 18.4.

⁸In cases where $N_{ab}^c > 1$ we would have had to keep track of an additional index μ at the a, b, c vertex. However, this index is also conserved around the loop meaning that the sum eventually becomes $\sum_{c,\mu} N_{ab}^c$ which will then generate a factor of N_{ab}^c as desired.

18.7 Appendix: Purely Algebraic Proof of Verlinde Relation

In this section we assume only that we have a set of symmetric fusion matrices $[N_a]_c^b$ which represent the fusion algebra⁹. Nowhere do we need to know anything about the braiding properties of the particle types (indeed, a braiding need not even be defined!). The fusion matrices must be commutative as in Eq. 8.9 so that they are all simultaneously diagonalizable by a unitary matrix which we will call U for now (See Eq. 8.11) which we write as

$$N_a = U \lambda^{(a)} U^\dagger \quad (18.7)$$

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

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

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

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

$$U_{ad} = \sum_c [N_a]_0^c U_{cd} = U_{0d} \lambda_d^{(a)}$$

so that

$$\lambda_d^{(a)} = U_{ad} / U_{0d}$$

substituting back into Eq. 18.7 we get

$$[N_a]_b^c = \sum_x U_{bx} \frac{U_{ax}}{U_{0x}} U_{cx}^*. \quad (18.8)$$

⁹The argument of this section is reproduced from Bonderson, Patel, Shtengel, and Simon, to be published.

Using Eq. 8.6, we immediately obtain that

$$U_{cx} = U_{\bar{c}x}^*$$

In particular this implies U_{0x} is real. Using Eq. 12.1 (also the result of exercise 8.1)

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

we see that the vector \vec{d} must be an eigenvector of N_a for all a , and hence is a column of the matrix U . We choose this to be U_{0x} , and indeed this must be the eigenvector with the largest eigenvalue for each N_a .

Further reading

This is some reading

Exercises

Exercise 18.1 Kirby Color

From any anyon theory (i.e., TQFT or modular tensor category) we can construct a type of string (a sum of particle types) called an Ω (sometimes ω) string, or sometimes called a Kirby-color string as given in Eq. 18.6.

(a) Evaluating a knot diagram with the evaluation rules of the TQFT gives

$$Z(S^3 \text{ with link}) / Z(S^3)$$

So the empty diagram is give value 1.

Consider a simple ring (an “unknot” or unknotted loop of string), black-board framed (meaning no twists) of Kirby color string. Evaluate this diagram.

(b) A knot (or link) of Kirby-colored string is meant to be equivalent to doing surgery on a the knot thickened into a torus. Considering the result of part (a) above as well as part (a) of the above exercise 19.1 on Surgery. Are these results consistent?

(c) Show that the Ω string made into a loop has the so-called “killing property” shown in Fig. 18.5. In other words, any diagram gives zero unless the particle type going through the Ω loop is the trivial or vacuum particle. Hint: Use the fact that the quantum dimension is part of the modular S matrix, and various properties of the S matrix to prove this identity.

(d) Evaluate a Hopf Link of Kirby color string (See Fig. 19.11). Does this match the result of part (b) of the exercise 19.1 above?

(e) [Harder] Evaluate the Borromean rings of Kirby color string (See Fig. 19.12). Compare your result to that of part (c) of exercise 19.1 above, and also the discussion in the problem on “Ground State Degeneracy” above.

Hints: Consider the F-move shown in figure 18.11. By closing up the top

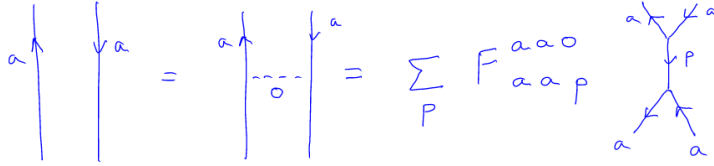


Fig. 18.11 An F-move. The far left diagram can be thought of as having a vacuum particle go between the two strings (middle). Then we can use an F-move to obtain the diagram on the right

and bottom, show that $F_{aa0}^{aa0} = 1/d_a$ with d_a the quantum dimension of the particle a . You will need the locality law (also known as the "no-tadpole" law), which says that diagrams of the type shown in Fig. 18.12 must be zero unless the incoming particle is the vacuum, $p = 0$.



Fig. 18.12 A tadpole diagram must be zero unless $p = 0$, by locality

Exercise 18.2 *Handle Slide* As discussed in section ??, one can describe a 3-manifold by giving a knot (or link) diagram which should be thickened into a tube and surgered. A handle-slide of a link diagram (which corresponds to sliding a handle of the manifold over another handle, but leaving the manifold topologically unchanged) involves splitting one strand, having it trace the path of a second strand and then reconnecting. An example is shown Fig. 19.7. In TQFT, one uses a string of Kirby color to represent the knot or link to be surgered. In fact, the evaluation of the link in the diagrammatic calculus is unchanged by handle-slides. While it takes a bit more diagrammatic calculus rules to derive the handle-slide invariance in general, a simple case of the handle-slide is fairly easy to derive. Consider instead a handle-slide over an untwisted loop as shown in the figure 18.13. Use the killing property. You will have to think about fusion, but you should not need to do any detailed calculations with F matrices.

Exercise 18.3 *Fusion and Ground State Degeneracy* In exercise 8.2 above, we have calculated the ground state degeneracy for a TQFT on an arbitrary oriented 2d manifold Σ . Using the Verlinde relation, show that the ground state degeneracy can be written as

$$\dim = Z(\Sigma \times S^1) = \sum_x [S_{0x}]^{\chi}$$

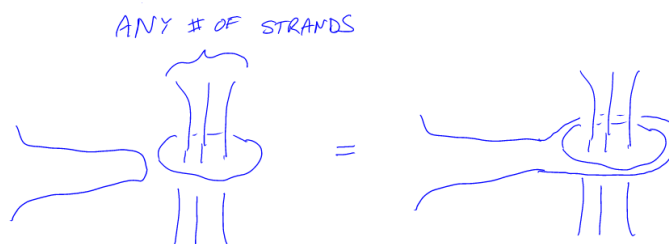


Fig. 18.13 A handle-slide over an untwisted loop. In this figure all strings are meant to be Ω strings or Kirby color (a weighted sum of all particle types).

where $\chi = 2 - 2g$ is the Euler characteristic where g is the *genus* of the manifold (the number of handles).

Surgery and More Complicated 3-Manifolds¹

19

In the previous chapter 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, but we will not need that here.¹) We would like to use this sort of trick to be able to study much more complicated three dimensional manifolds.

The understanding of three dimensional manifolds is a very rich and beautiful problem². In order to describe complicated manifolds it is useful to think in terms of so-called surgery. Similar to what we were just 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!⁴ By using successive surgeries we will be able to construct any orientable three-dimensional manifold.

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 .

²Many important results on three dimensional manifolds have been discovered recently. Perelman's³ 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!

³Grigori 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.

¹Although this chapter is super interesting and fun, physicists can probably skip it on a first reading.

¹See however, the discussion in section *** as well as ref *** for example.

⁴Prehensile toes could be useful I suppose!

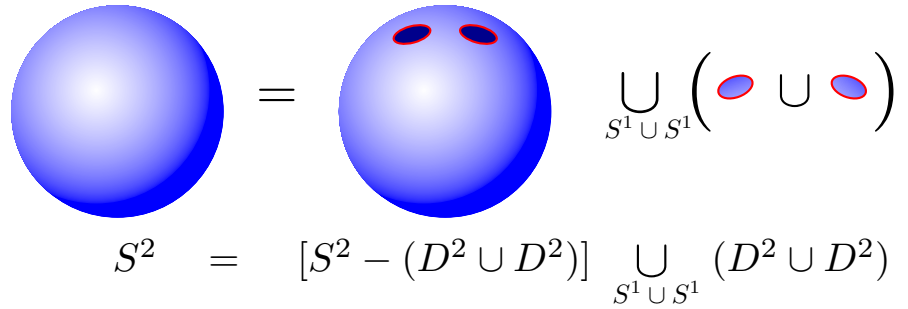
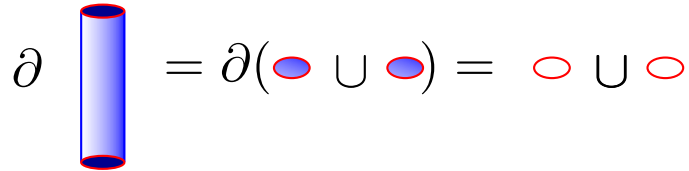


Fig. 19.1 Writing a sphere $\mathcal{M} = S^2$ as the union of two manifolds glued along their boundaries. \mathcal{M}_2 is the union of two disks $D^2 \cup D^2$. $\mathcal{M}_1 = S^2 - (D^2 \cup D^2)$ is the remainder. The two manifolds are glued along their common boundary $S^1 \cup S^1$.

19.1 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. 19.1. 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. 19.2.



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

Fig. 19.2 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. 19.3. The resulting manifold \mathcal{M}' is the torus T^2 .

Thus we have surgered a sphere and turned it into a torus. Note that there is another way to think of this procedure. If $\mathcal{M} = \partial\mathcal{N}$ then surgery on \mathcal{M} is the same as attaching a handle to \mathcal{N} . In the case we just considered we would take $\mathcal{N} = B^3$ the 3-ball (sometimes denoted

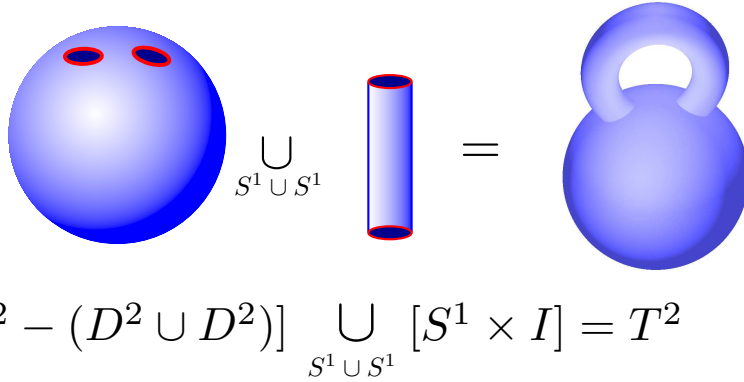


Fig. 19.3 Gluing the cylinder surface $\mathcal{M}'_2 = S^1 \times I$ to the manifold $\mathcal{M}_1 = S^2 - (D^2 \cup D^2)$ along their common boundary $S^1 \cup S^1$ gives the torus T^2 .

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. 19.4

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

Fig. 19.4 Handle attaching on the manifold \mathcal{N} is the same as surgery on a manifold $\mathcal{M} = \partial\mathcal{N}$.

19.2 Surgery on 3-manifolds

We can also perform surgery on three-dimensional manifolds.⁵ Start with a 3-manifold \mathcal{M} , such as perhaps the \mathbb{R}^3 space around us, or maybe S^3 . Now consider a solid torus

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

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

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

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

The idea of surgery is to remove \mathcal{M}_2 and replace it with \mathcal{M}'_2 to generate a new manifold \mathcal{M}' with no boundary.⁶ The reason this is difficult to visualize is because the new structure is not embeddable within the

⁵This is the part that is guaranteed to make your head explode.

⁶Stop here, think about what we have done. Collect the pieces of your exploded head.

original \mathbb{R}^3 . This is torus surgery on a 3-manifold, and it is called Dehn surgery. Another way to describe what we have done is that we have removed a solid torus, switched the meridian and longitude (switched the filled-contractable and the unfilled-uncontractable) and then glued it back in. In fact, one can make more complicated transformations on the torus before gluing it back in (and it is still called Dehn surgery) but we will not need this.

It is worth noting that the solid torus we removed could be embedded in a very complicated way within the original manifold — i.e., it could follow a complicated, even knotted, path, as in the figure on the right of Fig. 7.10.

19.2.1 Lickorish-Wallace Theorem

⁷In 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!

An important theorem⁷ 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 follow some knotted path).

⁸See appendix *** for a more precise definition of this prescription

One has the following procedure. We start with a link (some knot possibly of several strands), embedded in S^3 . Thicken each line to a solid torus. Excise each of these solid tori, and replace them by tori with longitude and meridian switched⁸. Any possible 3-manifold can be obtained in this way by surgering an appropriately chosen link. We summarize with

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

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

19.2.2 Kirby Calculus

It is not the case that all topologically different links, when surgered, give topologically different manifolds. Fortunately, the rules for which knots give the same manifolds have been worked out. These rules, known as Kirby calculus, are stated as a set of 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

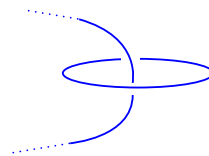


Fig. 19.5 A circumcision. Both strings can be removed. This is a third Kirby move which is implied by the first two only if you start with a knot embedded in S^3 . Otherwise it is an independent move that is required. See ***

transformations. Perhaps the simplest set of two elementary basic moves are known as Kirby moves:⁹

(1) **Blow up/ Blow Down:**¹⁰ One can add or remove an unlinked loop with a single twist, as shown in Fig. 19.6.

Addition or Removal of

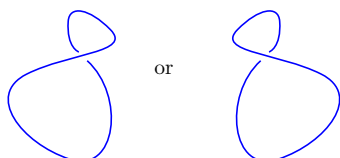


Fig. 19.6 Blow up/ Blow down. Addition or removal of an unlinked loop with a single twist leaves the 3-manifold represented by surgery on the knot unchanged.

(2) **Handle-Slide:**¹¹ A string can be broken open and pulled along the full path of another string, and then reconnected. See Fig. 19.7.

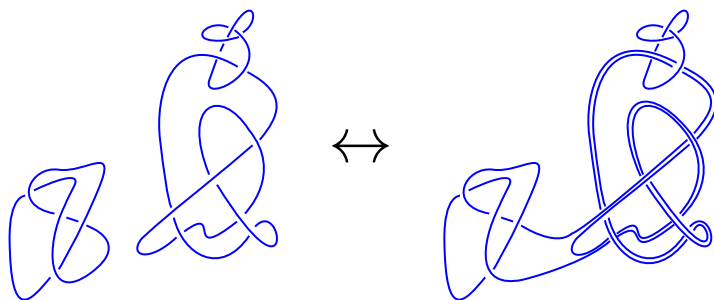


Fig. 19.7 A handle-slide move. (See Fig. 19.9 for another example.) Both left and right sides of this pictures represent the same 3-manifold after surgery. Note that the knot that is slid over (the right of the two pieces on the left) is 0-framed, meaning that it has self-twisting (writhe) of zero. If this were not the case the strand that slides over it would also twist around it. See *** for details.

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. While these moves may seem strange, it is possible to develop some geometric intuition for what these mean. This is addressed roughly in the appendix to this chapter.

19.2.3 Witten-Reshitikhin-Turaev Invariant

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

Our strategy of building a manifold invariant is to describe the manifold by using surgery on a link. Given knowledge of the rules of Kirby calculus, to construct a manifold invariant for three manifolds, we need

¹⁰The 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.

¹¹The nomenclature “handle slide” comes from an interpretation of this move as sliding handles around on a manifold. Consider the example used in section 19.1 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, which is what the handle-slide represents.

only construct a knot invariant that is invariant under Kirby moves. Being that the Chern-Simons path integral is not really well defined as a path integral, it turns out that this scheme is a way to make the manifold invariants of Chern-Simons theory mathematically rigorous [Reshetikhin and Turaev, 1991; Lickorish, 1993; Witten, 1989].

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

The first Kirby move (The blow up/blow down) does not sound so hard to finagle just by using some normalization factor for each twist and loop (We will show details of this later in ***). The second Kirby move seems harder to achieve, but can be achieved if one uses the so-called Kirby color combination (sometimes known as an Ω string)

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

where here we mean that we are summing over particle types a , and S is the modular S -matrix, and the subscript 0 refers to the vacuum or identity particle. Diagrammatically we have Fig. 19.8. It turns out (See

$$\text{red line} = \sum_a S_{0a} \text{blue line with arrow } a$$

Fig. 19.8 A String of Kirby color is a weighted superposition of all anyon string types. Note that the Kirby color string does not have an arrow on it since it is an equal sum over all pairs of particles and their antiparticles.

exercise 18.2) that the corresponding knot invariant that comes from evaluating a knot of Kirby color is invariant under handle-slides. The manifold invariant that results from evaluating the corresponding knot invariant of the Kirby-color string is known as the Witten-Reshetikhin-Turaev invariant and it gives a rigorous re-definition of the Chern-Simons manifold invariants defined by Witten. (See chapter *** for more details).

19.3 Appendix: Details of Surgery and Meaning of Kirby Moves

The point of this appendix is give a bit more detail of the surgery prescription and to give some geometric intuition for why Kirby moves on a link leave the resulting manifold described by the link unchanged.

First let us bit 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 ***). 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.

Blow Up / Blow Down: Consider the twisted loop in Fig. 19.6 embedded in S^3 . We would like to perform surgery on this loop and we claim we still obtain S^3 . As described in Fig. 2.7 a string with a small twist loop as in Fig. 19.6 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. 19.9 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. Since this line goes around each handle once, this results in precisely the construction of $L(1, 1) = S^3$ as described in 7.4.

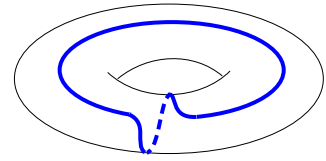


Fig. 19.9 A line that wraps both the longitude and meridian of the torus. If we thicken the knot shown in Fig. 19.6 to a torus and draw a line around the longitude of the torus, then try to straighten the torus out to remove the twist, the straight line ends up looking like this.

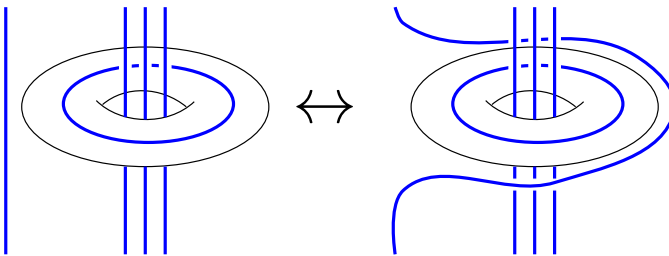


Fig. 19.10 An example of a simple handle-slide move.

Handle-Slide: It is fairly easy to describe why the handle-slide is an allowed move. Consider the simple handle-slide shown in Fig. 19.9. 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. This remains true even if the torus is embedded in the manifold in a complicated way, as in Fig. 19.7.

Further Reading

For more detailed discussion of Surgery and Kirby Calculus, as well as a nice discussion of manifold invariants, see

- V. V. Prasolov and A. B. Sossinsky, *Knots, Links, Braids and 3-Manifolds: An introduction to the New Invariants in Low-Dimensional Topology*, Translations of Mathematical Monographs, v 154, American Mathematical Society, Providence RI, (1996).

The following references are standards for Surgery and Kirby Calculus, although they emphasize four dimensional topology.

- Robert E. Gompf and András I. Stipsicz, *4-Manifolds and Kirby Calculus*, American Mathematical Society, Graduate Studies in Mathematics Volume: 20 (1999).
- Robion Kirby, *The Topology of 4-Manifolds*, Springer (1989).
- Selman Akbulut, *4 manifolds*, Oxford Graduate Texts in Mathematics (2016).

Exercises

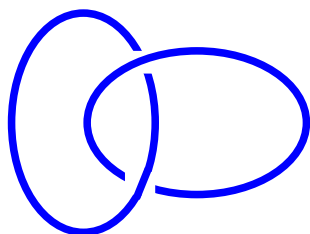


Fig. 19.11 Hopf Link

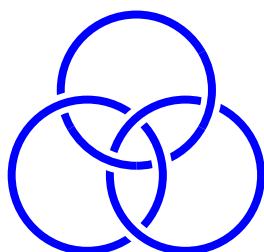


Fig. 19.12 Borromean Rings. Cutting any one strand disconnects the other two.

Exercise 19.1 Surgery

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

(b) [Not hard if you think about it right!] Consider two linked rings, known as the Hopf link (See Fig. 19.11). 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 .

(c) [Hard] Consider the link shown in Fig. 19.12 known as the Borromean rings¹². This is an interesting link because no two strands are actually linked with each other, but the three links are still tied together. If you remove any one strand the remaining two come apart.

Consider starting with S^3 and embedding the Borromean rings within the S^3 with “blackboard framing”. Thicken all three strands into solid tori and perform surgery on each of the three links exactly as we did above. What manifold do you obtain? Hint 1: Think about the group of topologically different loops through the manifold starting and ending at the same point, the so-called “fundamental group” or first homotopy group. (See section 28.3). Hint 2: If we say a path around the meridian of one of the three Borromean

¹² 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 “Wal-knot”, or “the knot of the slain.”

rings (i.e., threading through 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 topologically 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.)

Quantum Error Correction and The Toric Code

20

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.

20.1 Classical Versus Quantum Information

20.1.1 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 qubit² 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

²Sometimes q-bit, but never cubit.

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$$

with arbitrary complex prefactor α, β (where we normalize wavefunctions so $|\alpha|^2 + |\beta|^2 = 1$).

A quantum memory with N qubits is a vector within the 2^N dimensional complex Hilbert space. So for example, with 2 qubits the general state of a system is specified by four complex parameters

$$|\psi\rangle = \alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle \quad (20.1)$$

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!

20.1.2 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 20.1 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 occurred. If an error does occur on one of the physical bits (i.e., if one of the bits is accidentally flipped) 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 mechanical 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.

20.2 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 systems³!

As with so many great ideas in this field, the Toric code was invented by Kitaev (Kitaev 1997).

20.2.1 Toric Code Hilbert Space

We imagine an N_x by N_y square lattice with spins on each edge, where the edges of the lattice are made periodic hence forming a torus (hence the name “toric”). The total number of spins is $N = 2N_xN_y$ and correspondingly the dimension of the Hilbert space is 2^N .

³The 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?

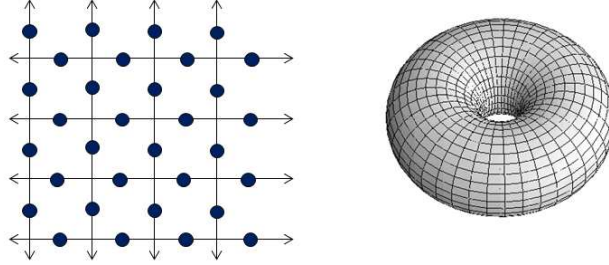


Fig. 20.1 The Hilbert space of the toric code — an N_x by N_y square lattice with spins (dots) on each edge wrapped up to make it periodic in both directions — i.e., a torus. Hence the name. There are 32 spins in this picture so the Hilbert space has dimension 2^{32} .

⁴Caution: 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.

We will work with a basis in our Hilbert space of up and down spins⁴. A convenient notation is then to color in the edges containing down spins but leave uncolored the edges with up spins. See Fig. 20.2.

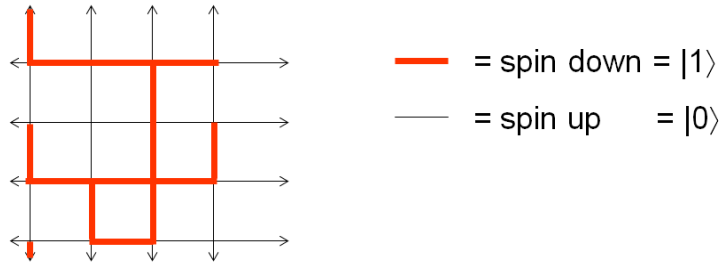


Fig. 20.2 A particular basis state of the Hilbert space, working in the up-down basis (z-eigenstates). Here we denote down spins by thick (red) lines. And up spins are denoted by not coloring in the edges.

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.

20.2.2 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

$$V_\alpha = \prod_{i \in \text{vertex } \alpha} \sigma_i^z$$

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. 20.3. Note that there are a total of $N_x N_y$ vertex operators.

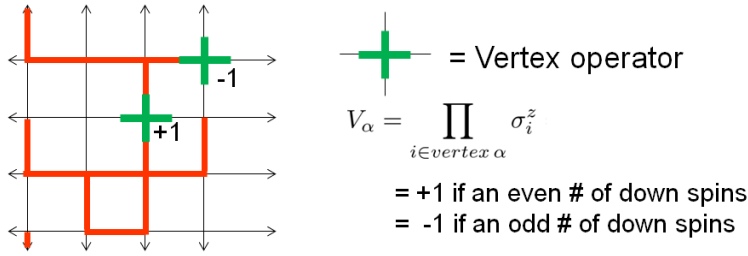


Fig. 20.3 The vertex operator returns $+1$ if there are an even number of incident down spins at that vertex and returns -1 if there are an odd number.

Note that it is possible (and useful) to define a corresponding projection operator

$$\tilde{V}_\alpha = \frac{1}{2}(1 + V_\alpha) \quad (20.2)$$

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

$$P_\beta = \prod_{i \in \text{plaquette } \beta} \sigma_i^x$$

which flips the state of the spins on all of the edges of the plaquette as depicted in Fig. 20.4. There are a total of $N_x N_y$ plaquette operators.

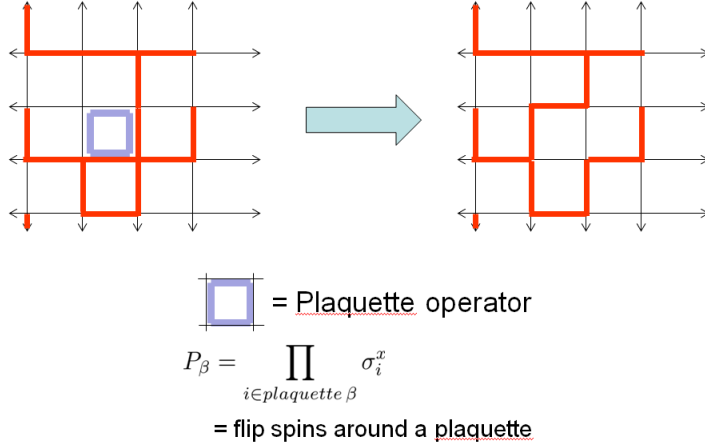


Fig. 20.4 The plaquette operator flips the state of the spin on the four edges of a plaquette.

As with the vertex operator, $P_\beta^2 = 1$ meaning P_β has eigenvalues $+1$ and -1 . We can similarly define a projector

$$\tilde{P}_\beta = \frac{1}{2}(1 - P_\beta) \quad (20.3)$$

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 plaquette. As such, the 0 eigenstate of \tilde{P}_β 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. 20.5. The orthogonal superposition (adding the two states with a - sign) will give the other eigenstate.

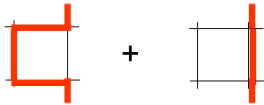


Fig. 20.5 A linear superposition of a flipped and unflipped plaquette is a $+1$ eigenstate of P_β or equivalently a 0 eigenstate of \tilde{P}_β . The -1 eigenstate is given by the orthogonal superposition, i.e., the superposition with a - sign between the two terms.

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 operators, 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 decomposition 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

$$\begin{aligned} \text{Number of Vertex Ops} + \text{Number of Plaquette Ops} - 2 + 2g \\ = \text{Number of Spins} \end{aligned}$$

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.

20.2.3 Building the code space

We are going to state two rules for constructing our code. We are imagining here that we have a great deal of control over the spins (the microscopic qubits) making up our system and we can impose these rules by fiat.

Rule 1: Specify that $V_\alpha = 1$ for every vertex (or equivalently $\tilde{V}_\alpha = 0$).

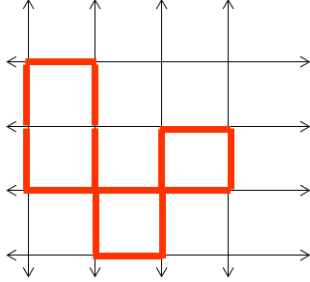


Fig. 20.6 A loop configuration consistent with the constraint that $V_\alpha = 1$ on every vertex. There must be an even number of red lines incident on every vertex.

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. 20.6

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. 20.5 this assures that every plaquette is in an equal superposition 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 *stabilizers* 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 wavefunction that satisfies both Rule 1 and Rule 2: First start in any state of spins up and spins down which satisfies rule 1, i.e., is a loop configuration. Then add to this in a superposition every configuration that can

be obtained by flipping plaquettes. We thus have

$$|\psi\rangle = \sum_{\substack{\text{all loop configs that can} \\ \text{be obtained by flipping pla-} \\ \text{quettes from a reference} \\ \text{loop config}}} |\text{loop config}\rangle \quad (20.4)$$

By adding up all such configurations, we assure that every plaquette is in the correct superposition of flipped and upflipped and we satisfy Rule 2.

The key question is whether one can obtain all loop configurations by starting in a reference configuration 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. 20.7. If one flips a plaquette (blue in the fig) along this cut (green in 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. 20.4, 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. 20.8

$$\begin{aligned} |\psi_{ee}\rangle &= \left| \text{torus with 0 red loops} \right\rangle + \left| \text{torus with 2 red loops} \right\rangle + \left| \text{torus with 4 red loops} \right\rangle + \dots \\ |\psi_{eo}\rangle &= \left| \text{torus with 1 red loop around handle 1} \right\rangle + \left| \text{torus with 3 red loops around handle 1} \right\rangle + \dots \end{aligned}$$

Fig. 20.8 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.

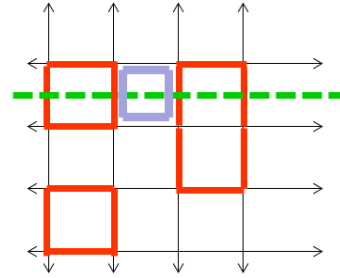


Fig. 20.7 Making a cut around one of the handles of torus, one can see that flipping a plaquette, such as the blue one, does not change the parity of the number of red bonds cutting the green line. Further, it does not matter where (at which y-coordinate) the green cut is made, the number of red bonds it cuts is always even.

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

$$|\psi\rangle = A_{ee}|\psi_{ee}\rangle + A_{eo}|\psi_{eo}\rangle + A_{oe}|\psi_{oe}\rangle + A_{oo}|\psi_{oo}\rangle \quad (20.5)$$

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. 20.1). We will refer to wavefunctions of the form of Eq. 20.5 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.

20.3 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.

20.3.1 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 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. 20.9.

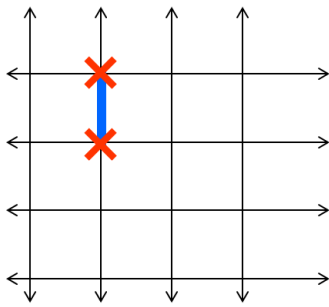


Fig. 20.9 A σ^x operator applied to the bond creates two vertices in the $V_\alpha = -1$ eigenstate.

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, returning 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. 20.10, 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.

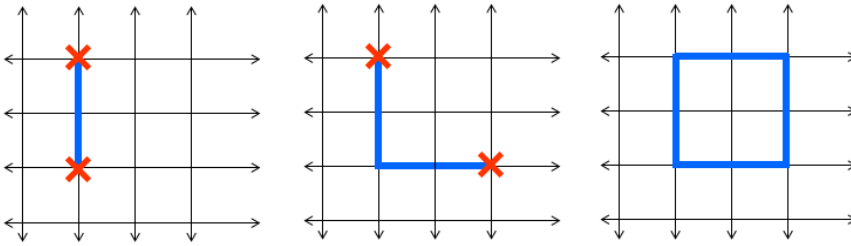


Fig. 20.10 Left: When two σ^x errors are made on bonds that share a vertex, the shared vertex is hit with σ_i^x twice, and thus becomes $V = +1$ again. Only the two vertices at the end of the "string" are in the $V = -1$ state. **Middle:** A longer string of errors. Note that we can only measure the 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 20.10 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. 20.10, 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. 20.10. 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 the 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 :

$$\begin{aligned} A_{ee}|\psi_{ee}\rangle + A_{eo}|\psi_{eo}\rangle + A_{oe}|\psi_{oe}\rangle + A_{oo}|\psi_{oo}\rangle \longrightarrow \\ A_{oe}|\psi_{ee}\rangle + A_{oo}|\psi_{eo}\rangle + A_{ee}|\psi_{oe}\rangle + A_{eo}|\psi_{oo}\rangle \end{aligned}$$

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.

20.3.2 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. 20.11. Recall that this eigenstate of the plaquette operator is a superposition of the flipped and unflipped plaquettes similar to that shown in Fig. 20.5 but with a minus sign between the two terms.

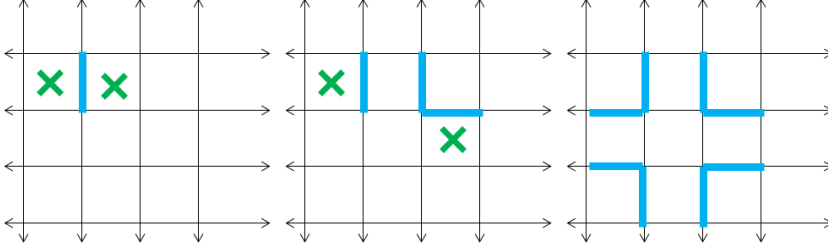


Fig. 20.11 **Left:** When a σ^z error is applied to a bond, the plaquettes on either side end up in the $P = -1$ state **Middle:** A string of several σ^z errors. **Right** A closed loop of σ^z errors. This is equal to the product of all of the enclosed V_α operators. In the code space, this is equal 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. 20.11. 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. 20.11, this loop of σ^z operators is equal to the product of the enclosed V_α operators. Since within the code space, $V_\alpha = 1$, a closed loop returns the system its original state. Another way of seeing this is to think in terms of the red loops of down spins discussed above. The σ_z operators register -1 each time they intersect a red loop. On the other hand the red loops must be closed so the number of intersections between a red loop and a closed 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. 20.12, 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

$$\begin{aligned} &A_{ee}|\psi_{ee}\rangle + A_{eo}|\psi_{eo}\rangle + A_{oe}|\psi_{oe}\rangle + A_{oo}|\psi_{oo}\rangle \longrightarrow \\ &A_{ee}|\psi_{ee}\rangle + A_{eo}|\psi_{eo}\rangle - A_{oe}|\psi_{oe}\rangle - A_{oo}|\psi_{oo}\rangle \end{aligned}$$

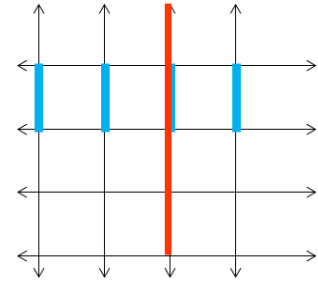


Fig. 20.12 If a string of σ^z goes around a handle, it measures the parity of the number of red strings going around the dual handle.

20.3.3 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!

20.3.4 More Comments on Errors

(1) A key point to take away here is that the *only* process which can cause logical errors 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 quite 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.

20.4 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 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

$$H = - \sum_{\text{vertices } \alpha} V_\alpha - \sum_{\text{plaquettes } \beta} P_\beta \quad (20.6)$$

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}_α and \tilde{P}_β defined by Eqs. 20.2 and 20.3. Writing

$$\tilde{H} = \sum_{\text{vertices } \alpha} \tilde{V}_\alpha + \sum_{\text{plaquettes } \beta} \tilde{P}_\beta \quad (20.7)$$

which differs from Eq. 20.6 only by 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 excitation has unit energy.

20.4.1 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 annihilated 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.

\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. 20.13 Fusion Table for the Toric Code

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. ??.

20.4.2 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 show 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 circle⁵.

⁵The experiment just described, while quite clear only tells us that e is either a boson or a fermion (since a fermion taken in a loop all the way around another fermion also accumulates no phase since it is equivalent to two exchanges).

To determine the phase of an exchange, we are going to attempt to do a twist in a world line as in Fig. 2.6 or 17.1. Considering Fig. 20.14

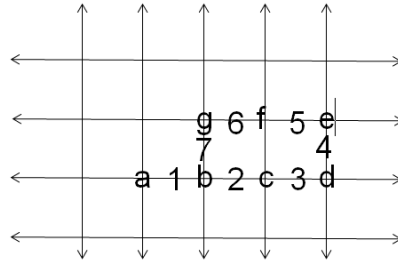


Fig. 20.14 Vertices are labeled with letters and bonds are labeled with numbers.

Now suppose there is initially an e particle at position a . One experiment we can do is to apply (reading right to left) $\sigma_1^x \sigma_2^x \sigma_3^x \sigma_4^x \sigma_5^x \sigma_6^x \sigma_7^x$. 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 positions c and d , moves the particle at d in a loop ($bgfe$) around c and annihilates 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 17.1. However, since the σ_x operators all commute, it must also be equal to the previously described process which just moves one particle around in a loop without introducing any twist. Hence we conclude that the e particle is a boson.

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. 20.10 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. 20.11 and is the product of a string of σ^z operators. Recall that the reason this process does not accumulate a phase is because the string of σ^z operators around the loop is equivalent to the product of the V_α vertex operators enclosed — and in the ground state, the V_α operators are in the $+1$ state. However, if there is one e particle inside the loop, this means that one of the V_α operators is actually in the -1 state. In this case the phase of taking the m particle around in a loop is actually -1 . So there is a phase of -1 for taking m around e .

Properties of f , the fermion

Since f is made up of an m bound to an e , it is easy to see that taking e around f accumulates a phase of -1 and taking m around f also accumulates a phase 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. 20.15

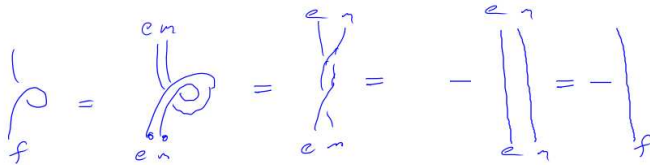


Fig. 20.15 The $f = e \times m$ particle is a fermion, since e braiding around m gives a -1 sign.

Note that taking f all the way around f will result in a net $+$ sign.

20.4.3 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 $\mathcal{D} = 2$.

20.4.4 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.

20.5 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 excitations 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 obtain⁶

$$|\tilde{\psi}\rangle = |\psi\rangle + (G\delta H)|\psi\rangle + (G\delta H)^2|\psi\rangle + \dots$$

⁶This is a Brillouin-Wigner perturbation theory, where successive terms are rigorously λ/Δ smaller.

and the energy modified by the perturbing Hamiltonian is then

$$E = \langle \tilde{\psi} | H_{\text{toric}} + \delta H | \tilde{\psi} \rangle$$

where here G is the greens function, which includes an energy denominator at least as big as the excitation gap Δ , so that successive terms in the expansion are smaller by order λ/Δ . The point here is that at M^{th} order in perturbation theory, we can only generate wavefunctions that differ from the original ground state by M applications of δH . Now recall that one cannot even distinguish the ground state sectors from each other unless one has a string operator that wraps all the way around the torus. Thus, the result of this calculation is identical for the four ground states out to very high order of perturbation theory, and any splitting of the four ground state sectors (or any mixing of the sectors) will be suppressed exponentially as $(\lambda/\Delta)^L$ which can be made arbitrarily small for a big system. It is clear that this general argument is not specific to the particular form of δH we have chosen.

One can go further and ask what happens to the excited particles when a perturbation is applied to the system. Similarly, we can perform a perturbation series. Here what happens is that the particles — which started as point defects — develop a nonzero length scale. As one moves a distance x further away from the particle, the influence of the presence of that particle decays as $(\lambda/\Delta)^x$. Again, if λ is small, then from a sufficiently far distance away, the particle again looks like a point. In particular, if one particle is braided around another at a sufficient distance away, it accumulates the expected phase that the pure toric code would have predicted. There are several strong arguments for 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 17.4 — it is not possible that the braiding phases in a theory change an arbitrarily small amount.

20.6 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 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

21

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 elements¹.

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. 21.1

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 $g \rightarrow g^{-1}$ as shown in Fig. 21.2.

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. 21.3. This is the string-net vertex fusion rule.²

¹I present this model on the “dual” graph compared to Kitaev’s presentation.

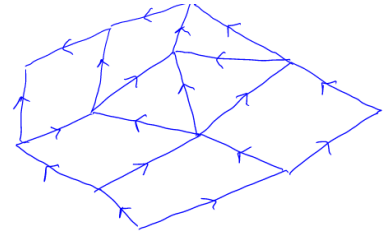


Fig. 21.1 Part of a directed graph.



Fig. 21.2 Inverting the direction on an edge inverts the group element.

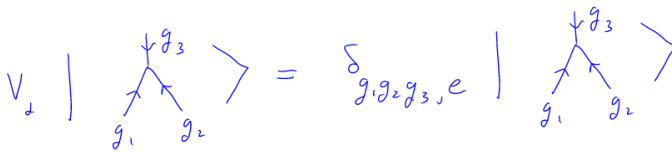


Fig. 21.3 Definition of V_α when all arrows are directed into the vertex (if a vertex is directed out, one can invert the arrow and invert the group element). The vertex operator gives zero unless the product of group elements around the plaquette gives the identity element e

²Note that if we have three edges coming into a vertex labeled g_1, g_2, g_3 the condition $g_1g_2g_3 = e$ is equivalent to $g_2g_3g_1 = e$ and $g_3g_1g_2 = e$.

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. 21.4.

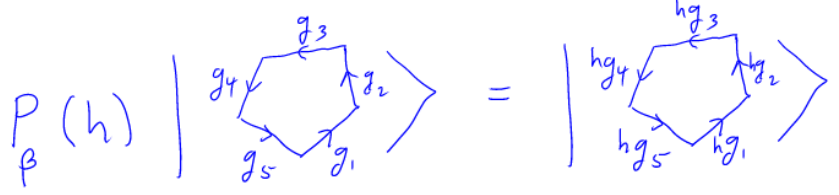


Fig. 21.4 The plaquette operator $P_\beta(h)$ premultiplies all of the clockwise oriented bonds by the element h .

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.

$$\begin{aligned} P_\beta(1) &= \text{identity operator} \\ P_\beta(-1) &= \text{multiply all edges by } -1. \text{ (i.e. flip all edges)} \end{aligned}$$

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 class³. 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. Essentially the wavefunction is given by a unique state plus everything that

³Two 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 $ug = gu$.

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.

21.0.1 \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 magnetic 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 corresponding \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 elementary 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 $0, \dots, (N-1)$, 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 nontrivial⁴: here it gives a phase of $e^{2\pi i/N}$.

⁴One 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.

21.1 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. 21.5. 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

$$ab = ba = c^{-1}$$

and in particular a and b must commute.

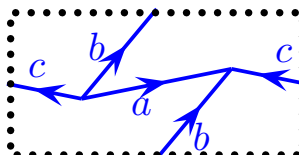


Fig. 21.5 The simplest possible decomposition of a torus into a single plaquette with three edges and two trivalent vertices. Here the dotted lines are periodic boundary conditions on the torus. The edges are labeled with particle types. The vertex conditions require $abc = I$ and $a^{-1}b^{-1}c^{-1} = I$.

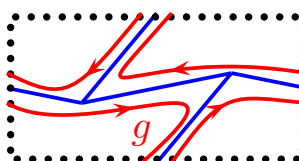


Fig. 21.6 The simplest possible decomposition of a torus into a single plaquette with three edges and two trivalent vertices. Here the dotted lines are periodic boundary conditions on the torus. The edges are labeled with particle types. The vertex conditions require $abc = I$ and $a^{-1}b^{-1}c^{-1} = I$.

Now let us think about the plaquette operator. In this lattice there is only a single

Use Burnside's lemma (otherwise known as the lemma not by Burnside)

More Generalizing The Toric Code: Loop Gases and String Nets

22

The general ideas presented with the toric code can be further generalized topologically ordered phases of matter. The 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.

22.1 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. 22.1 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

$$|\psi\rangle = \sum_{\substack{\text{all loop configs that can be} \\ \text{obtained from a reference} \\ \text{loop config}}} |\text{loop config}\rangle \quad (22.1)$$

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. 22.1 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. 22.2. We have always allowed smooth deformations in our diagrammatic algebras.

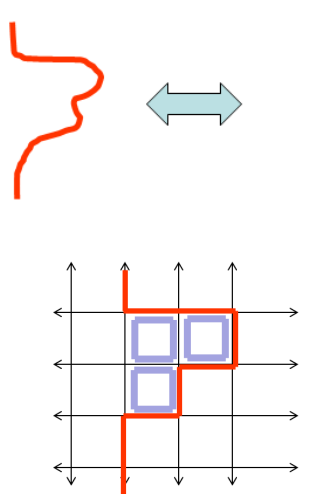


Fig. 22.2 Isotopy (Top) Off the lattice this is just deformation of a line. (Bottom) on the lattice, this is implemented by flipping over the blue plaquettes.

Move 2: "Adding or removing a loop". As shown in Fig. 22.3

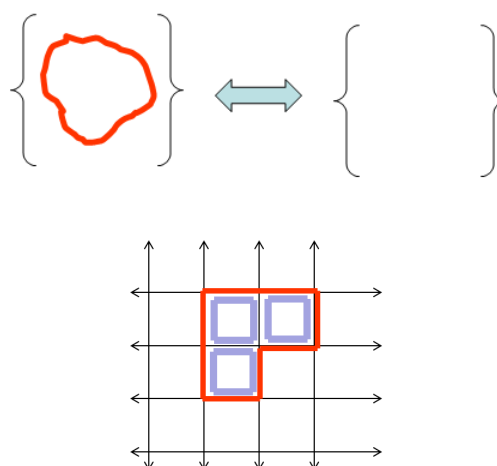


Fig. 22.3 Adding or Removing a loop (Top) Off the lattice (Bottom) On the lattice we flip the shown plaquettes.

Move 3: "Surgery" or reconnection of loops. As shown in Fig. 22.4

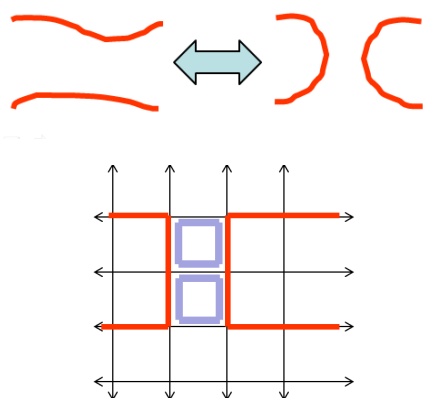


Fig. 22.4 Loop Surgery (Top) Off lattice surgery (Bottom) On lattice, flip the shown plaquettes

We can summarize these rules with simple skein-like relations as shown in Fig. 22.5

$$\bigcirc = 1$$

$$)(= \text{crossing}$$

Fig. 22.5 "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 wavefunction 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.

22.1.1 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. 22.6.



Fig. 22.6 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

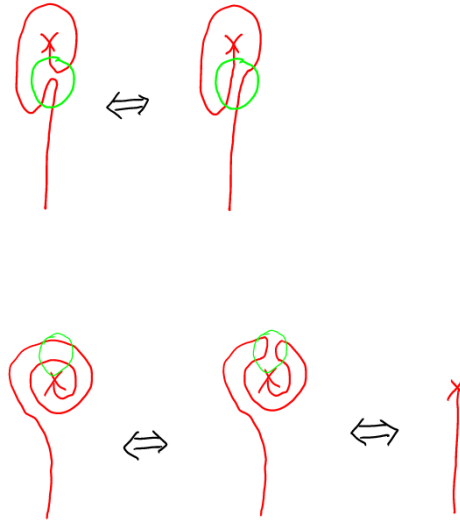


Fig. 22.7 Loop equivalences. Surgery is done inside the light green circles. The final equality on the lower right is just pulling the string tight.

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. 22.8

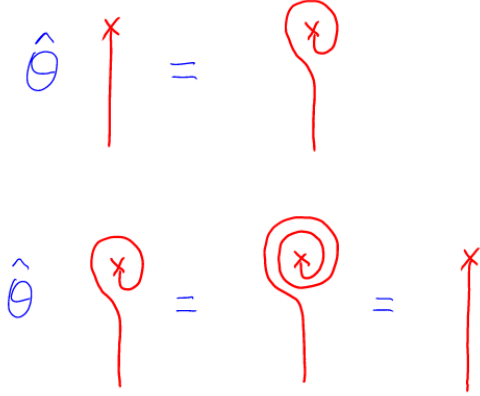


Fig. 22.8 Rotation

From these relations we can determine that the eigenvalues of the rotation operator are $+1$, corresponding to the e particle and -1 corresponding to the f particle, as shown in Fig. 22.9.

$$\hat{\theta} \left(\begin{array}{c} \times \\ \uparrow \end{array} \pm \begin{array}{c} \times \\ \text{loop} \\ \uparrow \end{array} \right) = \pm \left(\begin{array}{c} \times \\ \uparrow \end{array} \pm \begin{array}{c} \times \\ \text{loop} \\ \uparrow \end{array} \right)$$

Fig. 22.9 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. 22.10.

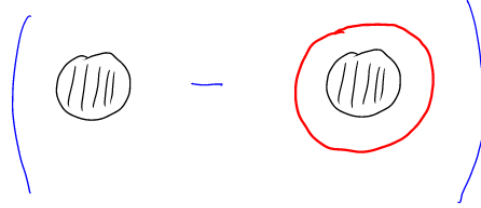


Fig. 22.10 The black disk is some region of our model. Forming a superposition of this region, and this region with a loop around it, with a minus sign between the two pieces, must be orthogonal to the ground state — it puts a magnetic excitation m in the region.

22.2 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.

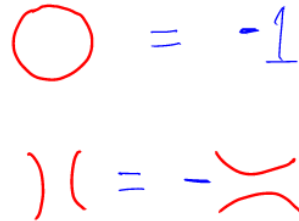


Fig. 22.11 "Skein" relations for the double-semion loop gas. Each loop removal/addition and each surgery incurs a minus sign. Note that these are the same as the Kauffman rules when we considered semions.

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

$$|\psi\rangle = \sum_{\substack{\text{all loop configs that can be} \\ \text{obtained from a reference} \\ \text{loop config}}} (-1)^{\text{Number of Loops}} |\text{loop config}\rangle \quad (22.2)$$

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.

22.2.1 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. 22.12). 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 condensed 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. 22.12. The advantage of this structure is that loops cannot intersect as they can (at the 4-fold corner) on the square lattice.

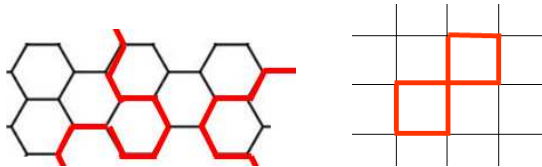


Fig. 22.12 Left: Toric code on a honeycomb, loops are nonintersecting. Right: On the square lattice loops can intersect at corners and one cannot tell if this picture represents one loop or two.

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. 22.13. Depending on the initial state, when the plaquette is flipped, one may or may not obtain a minus sign.

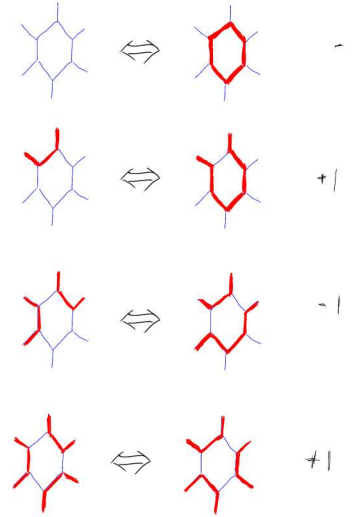


Fig. 22.13 Some plaquette flips for the double semion model on the hexagon. The top line obviously adds a loop, so should get a minus sign. The second line just stretches a loop over a plaquette, so does not get a minus sign. The third line is a surgery so gets a minus sign. The fourth line is a double 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

$$P'_\beta = \left(\prod_{i \in \text{plaquette } \beta} \sigma_i^x \right) (-1)^{\frac{1}{4} \sum_{j \in \text{legs of } \beta} (\sigma_j^z + 1)}$$

The overall Hamiltonian for this model is then

$$H = - \sum_{\text{vertices } \alpha} V_\alpha - \sum_{\text{plaquettes } \beta} P'_\beta$$

This Hamiltonian was first written down by Levin and Wen.

22.2.2 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. 22.14. Resulting in the effect of rotation being Fig. 22.15. Again we can use these to give us the eigenstates of the rotation operator as shown in Fig. 22.16

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

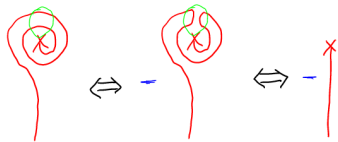


Fig. 22.14 Surgery incurs a minus sign. Compare to fig. 22.7

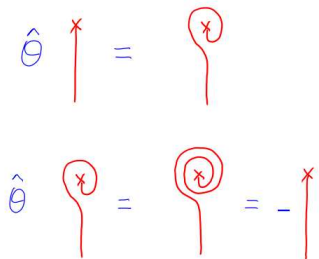


Fig. 22.15 Surgery incurs a minus sign. Compare to fig. 22.7

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!

$$\hat{\Theta} \left(\begin{array}{c} \text{red line with cross} \\ \text{blue loop with cross} \end{array} \right) = \pm i \left(\begin{array}{c} \text{red line with cross} \\ \text{blue loop with cross} \end{array} \right)$$

$$\hat{\Theta} \left(\begin{array}{c} \text{red line with cross} \\ \text{blue loop with cross} \end{array} \right) = \mp i \left(\begin{array}{c} \text{red line with cross} \\ \text{blue loop with cross} \end{array} \right)$$

Fig. 22.16 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 analogous to Fig. 22.10, 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. 22.17.

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 principle 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 as the positive eigenstate of the operator shown in Fig. 22.10 (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. 22.17 Fusion Table for Double Semion Model

¹To 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 18.4, and gives \mathcal{D} on the vacuum.

22.3 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. 22.18 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. 22.18. 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

$$|\psi\rangle = \sum_{\text{string nets}} \Phi(\text{net config}) |\text{net config}\rangle$$

where the prefactors $\Phi(\text{net config})$ satisfy some graphical rules as shown in Fig. 22.19.

$$\begin{aligned} \Phi \left(\begin{array}{c} \text{[grey]} \xrightarrow{i} \text{[grey]} \end{array} \right) &= \Phi \left(\begin{array}{c} \text{[grey]} \text{---} i \text{---} \text{[grey]} \end{array} \right) \\ \Phi \left(\begin{array}{c} \text{[grey]} \text{---} \text{[loop]}^i \end{array} \right) &= d_i \Phi \left(\begin{array}{c} \text{[grey]} \end{array} \right) \\ \Phi \left(\begin{array}{c} \text{[grey]} \xrightarrow{i} \text{[loop]}^k \xrightarrow{j} \text{[grey]} \end{array} \right) &= \delta_{ij} \Phi \left(\begin{array}{c} \text{[grey]} \xrightarrow{i} \text{[loop]}^k \xrightarrow{i} \text{[grey]} \end{array} \right) \\ \Phi \left(\begin{array}{c} \text{[grey]} \xrightarrow{i} \text{[loop]}^m \xrightarrow{j} \text{[grey]} \end{array} \right) &= \sum_n F_{kln}^{ijm} \Phi \left(\begin{array}{c} \text{[grey]} \xrightarrow{i} \text{[loop]}^l \xrightarrow{k} \text{[grey]} \end{array} \right) \end{aligned}$$

Fig. 22.19 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 changing 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 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 prefactor 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. 22.5 and Fig. 22.11.

It is important to note that the F -matrix used to define the string net (last line of Fig. 22.19) 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 resulting 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.

22.4 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. 22.20.

Starting with Eq. 9.1, we consider the following F -moves as shown in Fig. 22.21

$$\begin{aligned} \left. \begin{array}{c} \text{) } \end{array} \right(&= \phi^{-1} \begin{array}{c} \text{) } \\ \text{) } \end{array} + \phi^{-1/2} \begin{array}{c} \text{) } \\ \text{ | } \\ \text{) } \end{array} \\ \left. \begin{array}{c} \text{) } \end{array} \right(&= \phi^{-1/2} \begin{array}{c} \text{) } \\ \text{) } \end{array} - \phi^{-1} \begin{array}{c} \text{) } \\ \text{ | } \\ \text{) } \end{array} \end{aligned}$$

Fig. 22.21 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. 9.1.

We also expect to have rules of the form of Fig. 22.22 The first and second rules² are results of locality. The final rule is the usual rule that a



Fig. 22.20 A branching string net for the doubled Fibonacci model.

$$\begin{aligned} \bigcirc &= \chi \quad | \\ \bigcirc &= 0 \\ \bigcirc &= d \end{aligned}$$

Fig. 22.22 Rules for building the doubled fibonacci model.

²In 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. 22.21 to give Fig. 22.23.

$$\begin{aligned} \text{Red arc} \text{ } \bigcirc &= \phi^{-1} \text{Red arc} \text{ } \bigcirc \text{Red arc} + \phi^{-1/2} \text{Red arc} \text{ } \bigcirc \text{Red arc} \\ \text{Red arc} \text{ } \bigcirc &= \phi^{-1/2} \text{Red arc} \text{ } \bigcirc \text{Red arc} - \phi^{-1} \text{Red arc} \text{ } \bigcirc \text{Red arc} \end{aligned}$$

Fig. 22.23 Starting with Fig. 22.21 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. 22.21

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. 22.21 and 22.22. 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 22.22) 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. 22.21.

22.4.1 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. 22.24. Just as an example, consider the ending shown on the left of Fig. 22.25. By using an F -move, it is reduced to a combination of the three presented above.



Fig. 22.24 Possible string endings in the doubled fibonacci string net model.

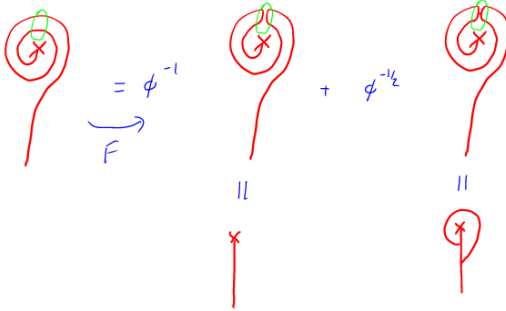


Fig. 22.25 An example of reducing a more complicated string ending into one of the three endings shown in Fig. 22.24.

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. 22.26 and then using F -matrices to reduce the result back to linear combinations of the same three possible endings.

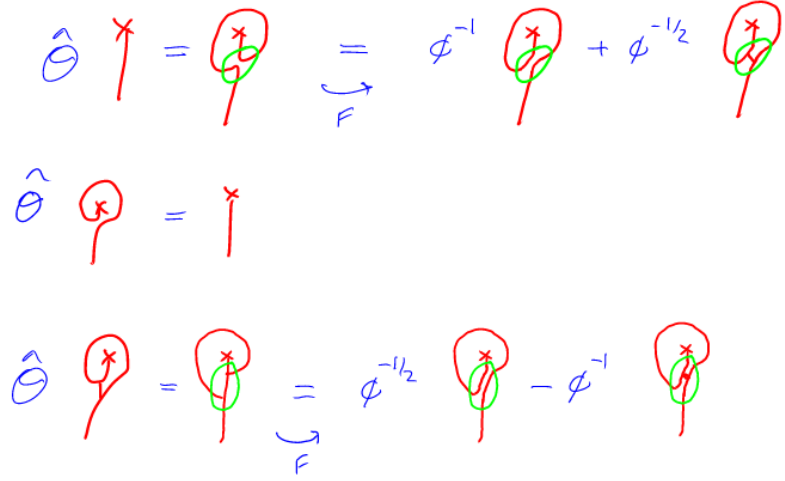


Fig. 22.26 The rotation operator $\hat{\Theta}$ applied to the possible string endings. Then using F matrices we reduce the results to linear combinations of the same endings.

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 factors 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 17.3.). 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!).

22.4.2 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 configuration – 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. 22.27

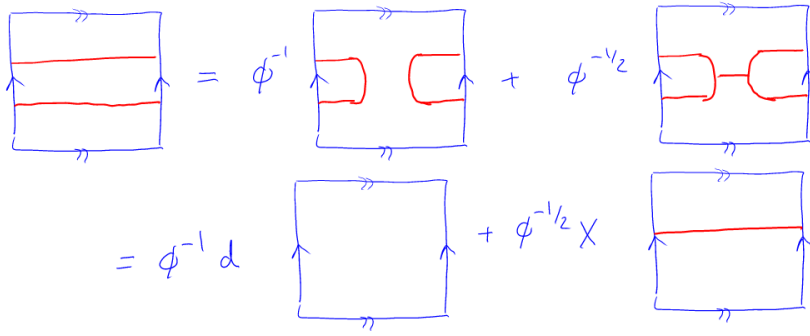


Fig. 22.27 Reducing two loops around a handle to a linear combination of one loops and zero loops.

22.5 Add details of Levin Wen Model on the lattice?

22.6 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. 22.28

$$\begin{aligned}
& \text{Diagram: two overlapping red loops with a green loop inside} \xrightarrow{F} \phi^{-1} \text{Diagram: two overlapping red loops with a green loop inside} + \phi^{-1/2} \text{Diagram: two overlapping red loops with a green loop inside} \\
& = \phi^{-1} (R_{zz}^I)^2 \text{Diagram: a red loop} + \phi^{-1/2} (R_{zz}^Z)^2 \text{Diagram: a red loop with a vertical line} \\
& = \left[\phi^{-1} (R_{zz}^I)^2 + \phi^{-1/2} (R_{zz}^Z)^2 \right] \text{Diagram: a red loop} \\
& = (R_{zz}^I)^2 + \phi (R_{zz}^Z)^2 = -1
\end{aligned}$$

Fig. 22.28 Calculating the nontrivial element of the Fibonacci anyon S -matrix.

Exercises

$$\begin{aligned}
\text{Diagram: a circle} &= 1 \\
\text{Diagram: two arcs meeting at a point} &=) (
\end{aligned}$$

Fig. 22.29 Loop gas rules for the toric code

Exercise 22.1 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. 22.29

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 22.30 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).

$$\begin{aligned}
\frac{1}{\sqrt{2}} \left(\text{Diagram: a vertical line with an 'x' at the top} + \text{Diagram: a vertical line with a circle and 'x' at the top} \right) &= \text{BOSON} \\
\frac{1}{\sqrt{2}} \left(\text{Diagram: a vertical line with an 'x' at the top} - \text{Diagram: a vertical line with a circle and 'x' at the top} \right) &= \text{FERMION}
\end{aligned}$$

Fig. 22.30 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. 22.31. 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. 22.31 Braiding defects

Exercise 22.2 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. 22.32. 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 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 22.33 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).

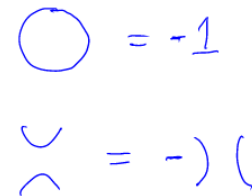


Fig. 22.32 Loop gas rules for the doubled semion model

$$\frac{1}{\sqrt{2}} \left(\begin{array}{c} \times \\ | \end{array} + i \begin{array}{c} \times \\ \bigcirc \end{array} \right) = \text{SEMION}$$

$$\frac{1}{\sqrt{2}} \left(\begin{array}{c} \times \\ | \end{array} - i \begin{array}{c} \times \\ \bigcirc \end{array} \right) = \text{SEMION}^*$$

Fig. 22.33 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 22.3 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. 22.34 (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. 22.35 for some values of the variables, d , X and T .

$$\begin{aligned}
 \left. \right) \left(&= \phi^{-1} \begin{array}{c} \text{---} \\ \text{---} \end{array} + \phi^{-1/2} \begin{array}{c} \text{---} \\ | \\ \text{---} \end{array} \\
 \left. \right) \text{---} \left(&= \phi^{-1/2} \begin{array}{c} \text{---} \\ \text{---} \end{array} - \phi^{-1} \begin{array}{c} \text{---} \\ | \\ \text{---} \end{array}
 \end{aligned}$$

Fig. 22.34 String net rules for the doubled Fibonacci model

$$\begin{aligned}
 \bigcirc &= d \\
 \bigcirc &= X \mid \\
 \vdash \bigcirc &= T \mid
 \end{aligned}$$

Fig. 22.35 Additional 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 22.36 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”.

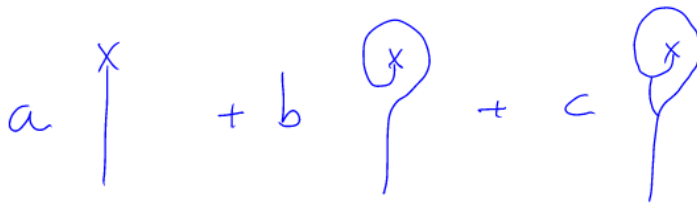


Fig. 22.36 Combination of defect types for the doubled Fibonacci model

Introduction to Quantum Hall — The Integer Effect

23

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 experiment¹. 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.

¹There are a good number of other contenders now. Probably the most convincing other case is ^3HeA 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.

23.1 Classical Hall Effect

In 1879 Edwin Hall discovered that when a current is run perpendicular to a magnetic field, a voltage is generated perpendicular to both field and current, and proportional to both (See Fig. 23.1). 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.

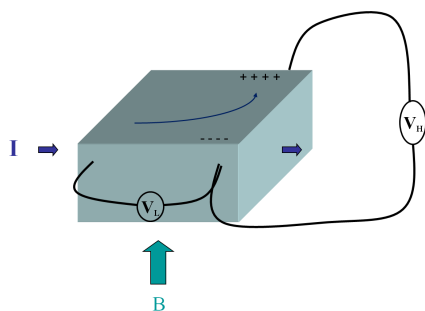


Fig. 23.1 Hall voltage V_H perpendicular to both magnetic field and current, and proportional to both. Also one measures a longitudinal voltage in the same direction as the current, roughly independent of magnetic field.

23.2 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

²Metal Oxide Semiconductor Field Effect Transistors

³More recently people have been able to produce materials like graphene which are literally one atom thick!

MOSFETs² and later quantum wells were used to provide a confining potential for electrons in one direction³, 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. 23.2.

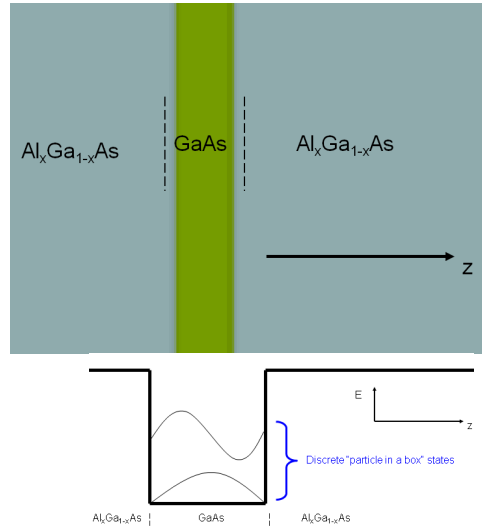


Fig. 23.2 **Top** A quantum well structure is a quasi-two-dimensional layer of one semiconductor sandwiched between two other semiconductors. **Bottom** The potential felt by an electron is like a particle in a box. If the energy is low enough, the electron is stuck in the lowest particle-in-box wavefunction $\varphi_0(z)$ 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. 23.2. 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).

23.3 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 mobility⁴ 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. 23.3.

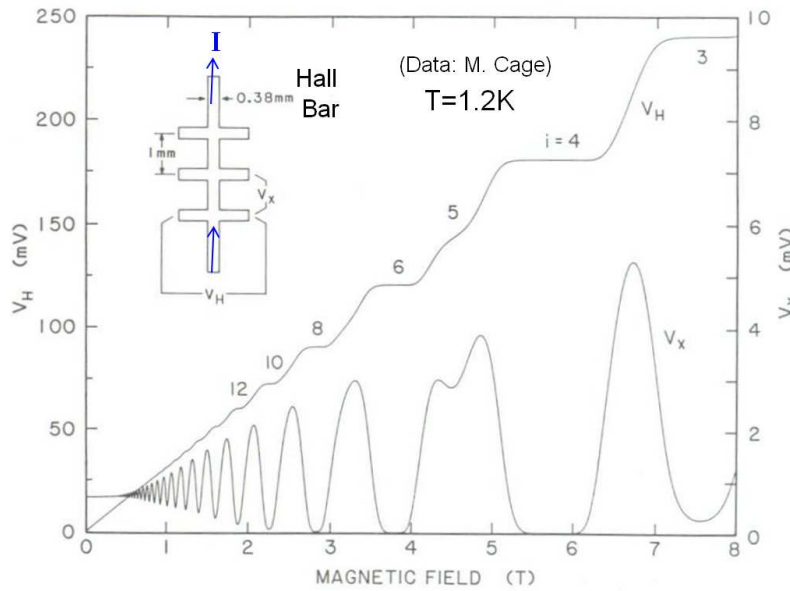


Fig. 23.3 An example of an Integer Quantum Hall experiment. The plateaus in V_H are such that $V_H = (1/i)(h/e^2)I$ with i the integer displayed over the plateau — where h is Planck's constant and e is the electron charge. At the same magnetic field where a plateau occurs in V_H the longitudinal voltage 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

$$V_H = \frac{1}{i} \frac{h}{e^2} I$$

⁴Meaning 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

$$R_H = \frac{1}{i} \frac{h}{e^2} = 1/G_H \quad (23.1)$$

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 resistance

$$R_L = 0$$

which implies we have a dissipationless state — similar to a superfluid. These statements become increasingly precise as the temperature is lowered.

⁵These are 2 by 2 matrices because they relate the vector electric field \mathbf{E} to the vector current \mathbf{j}

We should remember that conductivity and resistivities are both 2 by 2 matrices and are inverses of each other⁵. 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 longitudinal 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. 23.1 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.

⁶The 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.

23.4 Transport in Zero Disorder

⁷The 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.

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 $\hat{\mathbf{z}}$ -direction perpendicular to the plane. The Lorentz force on an electron will be

$$\mathbf{F} = -e(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

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.

23.5 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 open in the x direction, with length L_x (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

$$H_0 = \frac{1}{2m} ((p_x^2 + (p_y + eBx)^2))$$

where $p_j = -i\hbar\partial_j$.

The Hamiltonian is then translationally invariant 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

$$\left(\frac{p_x^2}{2m} + \frac{1}{2} m \omega_c^2 (k_y \ell^2 + x)^2 \right) \phi_{k_y}(x) = E \phi_{k_y}(x) \quad (23.2)$$

where ℓ is the so-called magnetic 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

$$E_p = \hbar\omega_c \left(p + \frac{1}{2} \right) \quad (23.3)$$

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. 23.4.

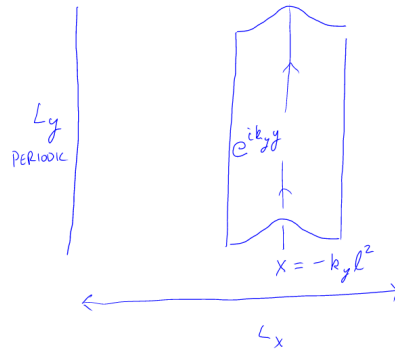


Fig. 23.4 The shape of the wavefunction of an electron in a magnetic field using Landau gauge. The form of the wavefunction will be harmonic oscillator on the x direction and plane-wave in the y -direction

Fixing the energy by fixing p in Eq. 23.3, the value of k_y is quantized in units of $2\pi/L_y$. Further, the position x ranges over L_x , meaning that k_y ranges over L_x/ℓ^2 . Thus the total number of possible values of k_y is

$$\text{Number of states in a Landau level} = \frac{L_x L_y}{2\pi\ell^2} = \frac{\text{Area}}{\phi_0} B$$

where

$$\phi_0 = h/e$$

is the magnetic flux quantum. Thus, the number of states in a Landau level is equal to the number of magnetic flux quanta of magnetic field incident on the plane.

We can plot the density of states for electrons in a magnetic field, as shown in Fig. 23.5

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

$$N_e = \nu N_\phi$$

Incompressibility 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 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

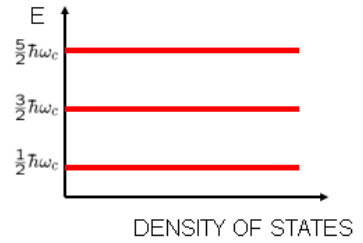


Fig. 23.5 The density of states for spin-polarized (or spinless) electrons in a magnetic field. At energies equal to half-odd integer multiples of the cyclotron frequency, there is a spike of degenerate states, with degeneracy $\frac{\text{Area}}{\phi_0} B$.

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. 23.6.

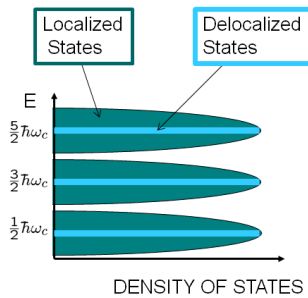


Fig. 23.6 The density of states for spin-polarized (or spinless) electrons in a magnetic field with disorder. The Landau bands are spread out, with localized eigenstates in the tails and extended eigenstates near the middle.

Since current tends to flow perpendicular to potential gradients (i.e., it is hall current), eigenstates tend to follow contours of constant potential. Thus many of the eigenstates at high and low energy will be trapped in local minima or maxima — isolated in a hill or valley and circling the peak or bottom. The result is that the eigenstates in the edge of the band experience localization, whereas (at least some) eigenstates near the center of the band as shown in Fig. 23.6.

When the chemical potential is anywhere in the localized states, then at low enough temperature, the electrons cannot move at all. Although there are states at this energy, they are all localized and electrons cannot jump between them. Hence we expect in this case that the DC dissipative conductance goes to zero. (For dissipative 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.

23.6 Laughlin's Quantization Argument

In 1981, shortly after von Klitzing's discovery of quantum Hall effect, Bob Laughlin⁸ 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.

⁸Laughlin would later go on to win a Nobel Prize for his explanation of *fractional* quantum Hall effect, which we will start discussing in chapter ***.

23.6.1 Byers and Yang Theorem

Consider any system (made of electrons and protons and neutrons) with a hole cut in it, as in Fig. 23.7. 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 Aharonov-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 Byers⁹-Yang¹⁰ theorem (1961).

To prove this theorem we use gauge invariance. One is always free to make a gauge transformation

$$\begin{aligned}\mathbf{A}'(\mathbf{r}) &= \mathbf{A}(\mathbf{r}) + (\hbar/e)\nabla\chi(\mathbf{r}) \\ \Psi'(\mathbf{r}_1, \dots, \mathbf{r}_N) &= \left[\prod_{j=1}^N e^{i\chi(\mathbf{r}_j)} \right] \Psi(\mathbf{r}_1, \dots, \mathbf{r}_N)\end{aligned}$$

which leave the physical electromagnetic field completely unchanged and changes the gauge of the wavefunction. The meaning of gauge invariance is that if we have a solution to the Schrodinger 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

$$\chi(\mathbf{r}) = m\theta(\mathbf{r})$$

where θ is the angle around the center. Making this gauge transform leaves the eigenspectrum of the system unchanged. However, the flux enclosed

$$\Phi' = \oint \mathbf{A}' \cdot d\mathbf{l} = \oint \mathbf{A} \cdot d\mathbf{l} + 2\pi m\hbar/e = \Phi + m\phi_0$$

has changed by an integer number of flux quanta.

23.6.2 Quantization of Hall Conductance

Laughlin's argument applies the Byers-Yang theorem to the Quantum Hall case. Consider a two dimensional electron system cut in an annulus¹¹ as shown in Fig. 23.8. 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 (dissipative) 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

$$\mathcal{E} = -\frac{d\Phi}{dt} = \oint d\mathbf{l} \cdot \mathbf{E}$$

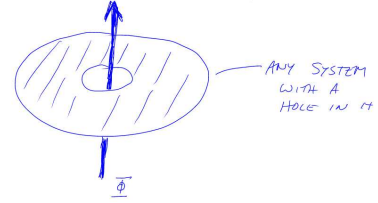


Fig. 23.7 The Byers-Yang theorem states that threading any integer number of flux quanta through a hole in a system leaves the eigenspectrum unchanged.

⁹Nina 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.

¹⁰Yang 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.

¹¹For studying current flow in magnetic fields, the annulus is known as "Corbino" geometry, after O. M. Corbino, who studied this in 1911.

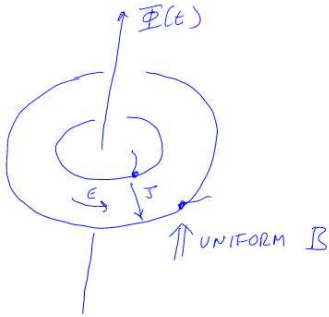


Fig. 23.8 Insertion of Flux $\Phi(t)$ through the center of an annulus of two-dimensional electrons in a uniform magnetic field. Adiabatically increasing the flux creates an electric field in the annular direction which then, by the Hall conductivity, creates current in the radial direction.

¹²There is a subtlety 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.

If there is a Hall conductance, G_H then this generates a radial current

$$J = G_H \mathcal{E}$$

As we slowly increase the flux by an amount $\Delta\Phi$ we have a total charge ΔQ moved from the inside to the outside of the annulus given by

$$\Delta Q = \int dt J(t) = G_H \int dt \mathcal{E}(t) = -G_H \int dt \frac{d\Phi(t)}{dt} = -G_H \Delta\Phi$$

Now the key to the argument is the Byers-Yang theorem. If we choose $\Delta\Phi = \phi_0$ a single flux quantum, then the final eigenstates of the system must be precisely the same as the initial eigenstates of the system. Since we have changed the system adiabatically (and there is a gap to excitations when the states at the chemical potential are localized due to disorder) the system must stay in the ground state¹² and the insertion of the flux quantum must take us from the ground state back to the very same ground state. The only thing that might have changed during this process is that an *integer* number p of electrons may have been transferred from the inside of the annulus to the outside. Thus we have

$$-pe = \Delta Q = -G_H \Delta\Phi = -G_H \phi_0 = -G_H (h/e)$$

Thus we obtain the quantized Hall conductance

$$G_H = p(e^2/h)$$

with p an integer!

Thus we see that the Hall conductance experiment is really some sort of “spectroscopy” to measure the charge on the electron! (hence the precision of the effect).

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 23.8 below.

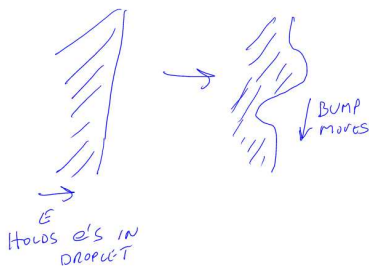


Fig. 23.9 A deformation of the edge is a low energy edge excitation which moves along the edge due to $E \times B$ drift.

23.7 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. 23.9. 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.

23.7.1 Landau Gauge Edge Picture for Integer Quantum Hall

Recall in Landau gauge (See section 23.5) 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. 23.10.

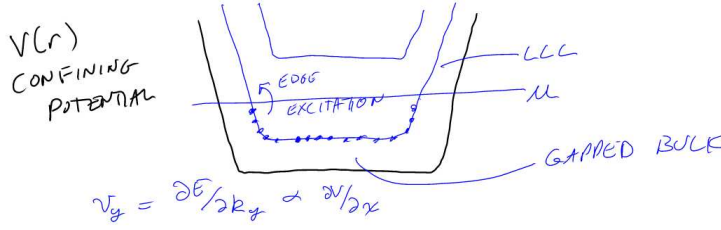


Fig. 23.10 Low energy edge excitations

The addition of the confining potential $V(x)$ simply adds this potential to the 1-d schrodinger equation 23.2. 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. 23.10.

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 empty 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

¹³The 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

$$\Delta n_{1d} = \frac{2\pi \Delta \mu}{\ell^2 \frac{\partial V}{\partial x}}$$

of fermions filled up to a chemical potential and they flow only in one direction along each edge — i.e., they are chiral fermions.

23.8 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 disorder. In particular we consider the geometry shown in Fig. 23.11.

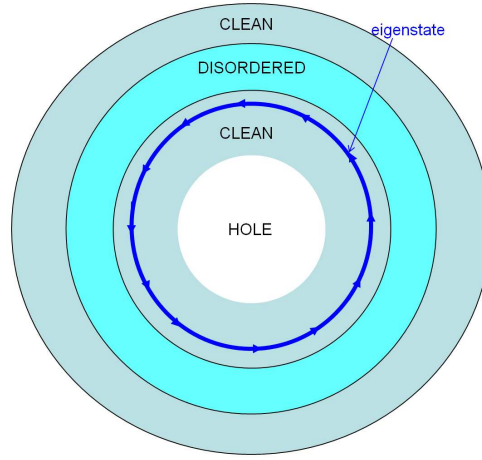


Fig. 23.11 The Halperin geometry. The same as the Laughlin annulus geometry, except here we add disorder only in part of the annulus. We have also shown (dark blue) a single particle eigenstate in the clean region, which forms a circle (with a small gaussian cross-section).

Here, the disorder is confined to only part of the annulus, the inner-most and outer-most regions of the annulus being disorder-free. Within the clean regions we can solve for the eigenstates using symmetric gauge (this is a homework problem, but we will also discuss further in the next chapter). The eigenstates are indexed by their angular momentum m , and in the Lowest Landau level, for example, they are given by

$$\varphi_m \sim z^m e^{-|z|^2/(4\ell^2)}$$

where $z = x + iy$ is the complex representation of the position. A radial cut of one of these eigenstates gives a gaussian wavepacket¹⁴ at radius

¹⁴Just find the maximum of $|\psi_m|^2$.

These then carry a net 1d electron current density

$$j = -ev\Delta n_{1d} = -e\left(\frac{1}{h}\frac{\partial V}{\partial x}\ell^2\right)\frac{2\pi\Delta\mu}{\ell^2\frac{\partial V}{\partial x}} = -(e/h)\Delta\mu$$

which is precisely the expected quantized Hall current flowing along the edge. ($\Delta\mu = -e\Delta V$).

$\ell\sqrt{2m}$ — very similar to what we had in Landau gauge, but now these eigenstates are indexed by angular momenta instead of linear momenta, and they go around in circle instead of going straight.

Let us imagine the chemical potential above the middle of a Landau level (say above the middle of the lowest Landau level) until it sits in a localized piece (at least within the disordered region the wavefunctions are localized). Since this is above the middle of the Landau level, the Landau level is completely filled in the clean region. 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 23.1 *Quantum Hall Conductivity vs Conductance*

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. 23.12. The conductivity tensor σ_{ij} relates the electric field to the current via

$$j_i = \sigma_{ij} E_j \quad (23.4)$$

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

$$\sigma = \begin{pmatrix} 0 & s \\ -s & 0 \end{pmatrix} \quad (23.5)$$

Show that the following two statements are true independent of the shape of the sample.

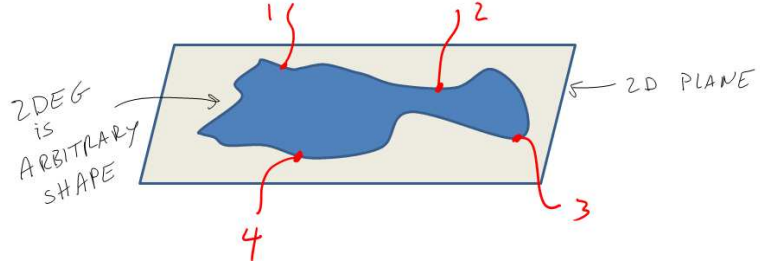


Fig. 23.12 A 2D electron gas of arbitrary shape with contacts 1,2,3,4 attached on its perimeter in clockwise order

(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 23.2 *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

$$H = \hbar\omega_c(a^\dagger a + \frac{1}{2}) \quad (23.6)$$

where $\omega_c = eB/m$ and

$$a = \sqrt{2}\ell \left(\bar{\partial} + \frac{1}{4\ell^2}z \right) \quad (23.7)$$

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

$$[a, a^\dagger] = 1 \quad (23.8)$$

and therefore that the energy spectrum is that of the harmonic oscillator

$$E_n = \hbar\omega_c(n + \frac{1}{2}) \quad (23.9)$$

(c) Once you obtain Eq. 23.6, show that any wavefunction

$$\psi = f(z)e^{-|z|^2/4\ell^2} \quad (23.10)$$

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

$$\psi_m = N_m z^m e^{-|z|^2/4\ell^2} \quad (23.11)$$

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

$$b = \sqrt{2}\ell \left(\partial + \frac{1}{4\ell^2} \bar{z} \right) \quad (23.12)$$

with $\partial = \partial/\partial z$, Show that the operator b also has canonical commutations

$$[b, b^\dagger] = 1 \quad (23.13)$$

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

$$L = \hat{z} \cdot (\vec{r} \times \vec{p}) = \hbar(b^\dagger b - a^\dagger a) \quad (23.14)$$

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

$$\pi_i^{(\alpha)} = p_i + \alpha \frac{\hbar}{2\ell^2} \epsilon_{ij} r_j \quad (23.15)$$

$$= \frac{\hbar}{\ell^2} \epsilon_{ij} \xi_j \quad (23.16)$$

defined for $\alpha = \pm 1$. Here $\alpha = +1$ gives the canonical momentum. Show that

$$[\pi_i^{(\alpha)}, \pi_j^{(\beta)}] = i\alpha\epsilon_{ij}\delta_{\alpha\beta} \frac{\hbar^2}{\ell^2} \quad (23.17)$$

The Hamiltonian

$$H = \frac{1}{2m}(p_i + eA_i)(p_i + eA_i) \quad (23.18)$$

can then be rewritten as

$$H = \frac{1}{2m} \pi_i^{(+1)} \pi_i^{(+1)} \quad (23.19)$$

with a sum on $i = \hat{x}, \hat{y}$ implied. Finally use

$$a = (-\pi_y^{(+1)} + i\pi_x^{(+1)}) \frac{\ell}{\sqrt{2}\hbar} \quad (23.20)$$

$$b = (\pi_y^{(-1)} + i\pi_x^{(-1)}) \frac{\ell}{\sqrt{2}\hbar} \quad (23.21)$$

to confirm that a and b are given by Eqs. 23.7 and 23.12 respectively. Finally confirm Eq. 23.6 by rewriting Eq. 23.19 using Eqs. 23.20 and 23.21.

A typical Place to get confused is the definition of ∂ . Note that

$$\partial z = \bar{\partial} \bar{z} = 1 \quad (23.22)$$

$$\bar{\partial} z = \partial \bar{z} = 0 \quad (23.23)$$

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

24

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

¹In this chapter we are thinking about non-interacting electrons in periodic potentials!

24.1 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 crystal². 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

²Some of the ideas discussed here do not depend too much on the system being precisely periodic.

$$|\Psi_{\mathbf{k}}^{\alpha}\rangle = e^{i\mathbf{k}\cdot\mathbf{r}}|u_{\mathbf{k}}^{\alpha}\rangle$$

where α is the band index, and $u_{\mathbf{k}}^{\alpha}(\mathbf{x})$ is a function periodic in the unit cell.

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

$$\mathbf{k} \rightarrow u_{\mathbf{k}}^{\alpha}(\mathbf{x}). \quad (24.1)$$

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 ,

$$e^{i\theta} \rightarrow e^{if(\theta)}$$

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. 24.1 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. 24.1 can be given in any dimension.

If we imagine continuously changing the physical Hamiltonian, this Chern number, which must be an integer, cannot change continuously. 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.

24.1.1 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.

24.2 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 mathematically³

$$2\pi(2 - 2g) = \int_M K dS$$

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

$$\mathcal{F}^\alpha(\mathbf{k}) = \epsilon^{ij} \langle \partial_{k_i} u_{\mathbf{k}}^\alpha | \partial_{k_j} u_{\mathbf{k}}^\alpha \rangle$$

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 by⁴

³The 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.

⁴The 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.

$$\sigma_{xy} = \frac{e^2}{h} \sum_{\text{filled bands } \alpha} C^\alpha$$

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 ***.

24.3 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 24.1, 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 deformed 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.

24.4 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

$$\langle j_a(x, t) \rangle = \frac{i}{\hbar} \int_{-\infty}^t dt' \int d\mathbf{x}' \langle [j_a(x, t), j_b(x', t')] A_b(x', t') \rangle$$

Introduction to Fractional Quantum Hall Effect

25

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 Stormer¹ 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 $\nu = 1/3$ plateau², but soon thereafter many more plateaus were discovered³. 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 through 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

¹Stormer 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.

²The 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!

³Over 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

$$\xrightarrow{\text{insert } \phi_0} |GS_1\rangle \xrightarrow{\text{insert } \phi_0} |GS_2\rangle \xrightarrow{\text{insert } \phi_0} |GS_3\rangle \xrightarrow{\text{insert } \phi_0} |GS_1\rangle \xrightarrow{\text{insert } \phi_0}$$

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 fractional 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 are $3N_e$ available single electron orbitals in which to place these electrons. Thus in the absence of disorder, and in the absence of interaction, there are

$$\binom{3N_e}{N_e} \sim (27/4)^{N_e}$$

multiparticle states to choose from — and all of these states have the same energy! In the thermodynamic limit this is an insanely enormous degeneracy⁴. 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.

⁴For 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.

25.0.1 Our Model Hamiltonian

Since we are to neglect disorder, we can write the Hamiltonian for our system of interacting electrons as

$$H = \sum_i \frac{(\mathbf{p}_i + e\mathbf{A}(\mathbf{r}_i))^2}{2m} + \sum_{i < j} V(\mathbf{r}_i - \mathbf{r}_j)$$

where the first term is just the kinetic energy of the electrons in the magnetic field, as discussed in Section 23.5, and the second term is the interaction between 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 with⁵.

Now we have already analyzed the first term in this Hamiltonian back in Eq. 23.5, 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 interaction 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

$$H = \sum_{i < j} V(\mathbf{r}_i - \mathbf{r}_j) \quad (25.1)$$

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!).

⁵For 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.

⁶Decades 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!

25.1 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 $\nabla \times \mathbf{A} = \mathbf{B}$.)

⁷We will ignore the spin degree of freedom as before.

In this gauge, lowest Landau level wavefunctions (as mentioned before in section 23.8) take the form⁷

$$\varphi_m(z) = C_m z^m e^{-|z|^2/(4\ell^2)} \quad (25.2)$$

where

$$z = x + iy = r e^{i\theta}$$

is the complex representation of the particle coordinate, $\ell = \sqrt{\hbar/eB}$ is the magnetic length, C_m is a normalization 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. 25.2. 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 magnetic flux.

25.1.1 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 have⁸

$$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = (\text{Polynomial in } z_1, \dots, z_N) \prod_{i=1}^N e^{-|z_i|^2/(4\ell^2)}$$

⁸The polynomial can also be chosen so as to have all real coefficients. This is because the Hamiltonian, once projected to a single Landau level, i.e., Eq. 25.1, 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, \dots, 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.

⁹While 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!

25.2 Laughlin's Ansatz

Laughlin simply guessed that a good wavefunction would be of the form¹⁰

$$\Psi_{\text{Laughlin}}^{(m)} = \prod_{i < j} (z_i - z_j)^m \prod_{i=1}^N e^{-|z_i|^2/(4\ell^2)}$$

¹⁰Note 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

$$\nu = N/N_\phi \rightarrow 1/m \quad \text{in large } N \text{ limit}$$

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.

25.2.1 Exact statements about Laughlin Wavefunction

It turns out that the Laughlin wavefunction is actually the exact ground state of a special inter-particle interaction¹².

Bosons at $\nu = 1/2$

Consider a system of bosons with the interparticle interaction given by¹³

$$V = V_0 \sum_{i < j} \delta(\mathbf{r}_i - \mathbf{r}_j)$$

with $V_0 > 0$. This is a non-negative definite interaction.

It is clear that the $\nu = 1/2$ Laughlin state of bosons $\Psi_{\text{Laughlin}}^{(m=2)}$ 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 property¹⁴. For example, the Laughlin state times any polynomial is also a zero energy state of this interaction, but since it has been multiplied by a polynomial, the total degree of the wavefunction is higher, meaning the wavefunction extends to higher radius, making the system less dense. A schematic of the ground state energy as a function of filling fraction for this case is shown in Fig. 25.1.

¹¹Roughly the story is as follows. The probability $|\Psi(z_1, \dots, z_N)|$ of finding particles at position z_1, \dots, z_N can be phrased as a classical stat mech problem of a one-component 2d coulomb plasma in a background charge, by writing

$$|\Psi|^2 = e^{-\beta U(z_1, \dots, z_N)}$$

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 positive 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.

¹²This was discovered by Haldane in 1983, then again by Trugman and Kivelson and also Pokrovski and Talapov in 1985.

¹³Actually this is a very realistic interaction for cold atom bosonic quantum Hall effect, should it be produced in the future.

¹⁴Although with some thought this fact seems obvious, proving it rigorously is tricky.

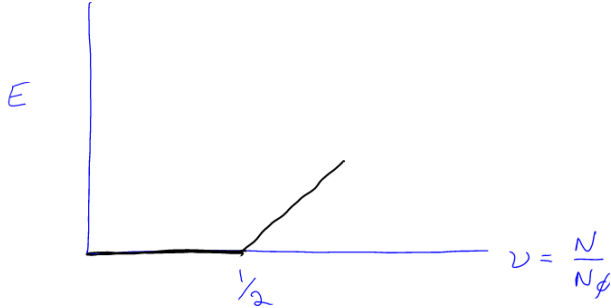


Fig. 25.1 Schematic of the ground state energy as a function of filling fraction for bosons with delta function interaction.

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 $\Psi(r_1, \dots, r_N)$ the interaction energy will be

$$E = \sum_{i < j} \int \mathbf{dr}_1 \dots \mathbf{dr}_N |\Psi|^2 \nabla^2 \delta(\mathbf{r}_i - \mathbf{r}_j)$$

Writing

$$\Psi(\mathbf{dr}_1 \dots \mathbf{dr}_N) = \phi(z_1 \dots z_N) \prod_{i=1}^N e^{-|z_i|^2/(4\ell^2)} \quad (25.3)$$

¹⁵Generally one would expect derivatives of the gaussian part as well when we integrate by parts. However, because the polynomial is antisymmetric, the derivative must act on the polynomial part to prevent the wavefunction from vanishing when particle coordinates coincide.

¹⁶Note that by antisymmetry the wavefunction must vanish as an odd number of powers as two particle positions approach each other.

with ϕ meaning 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 parts¹⁵ using $\nabla^2 = 4\partial_z\partial_{z^*}$ to give

$$E = \sum_{i < j} \int d\mathbf{r}_1 \dots d\mathbf{r}_N |\partial_{z_i}\phi|^2 \delta(\mathbf{r}_i - \mathbf{r}_j) \prod_{i=1}^N e^{-|z_i|^2/(2\ell^2)}$$

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 positive 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 analogously 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.

25.2.2 Real Interactions

¹⁷In 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. 25.2), the interaction gets modified – this mainly effects the short range behavior.

¹⁸The full Hilbert space is 45207 dimensional!

Obviously electrons do not interact via a $\nabla^2\delta$ interaction. They interact via a Coulomb interaction¹⁷ 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 space¹⁸ 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 thousand¹⁹.

25.3 Quasiparticles

The Laughlin quantum hall ground state is a uniform density fluid (we will actually show this as a homework problem). Density perturbations are made in discrete units of charge known as *quasiparticles*. Positively charged bumps of charge (opposite the charge of the electron) are known as *quasiholes* and negatively charged bumps of charge (same charge of the electron) are *quasielectrons*.

25.3.1 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

¹⁹I need to recheck this number***.

quasihole at position $\mathbf{0}$. 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 orbital open at position $\mathbf{0}$. 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 position $\mathbf{0}$. 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

$$\Psi_{qhs}(w_1, \dots, w_M) = \left[\prod_{\alpha=1}^M \prod_{i=1}^N (z_i - w_\alpha) \right] \Psi_{Laughlin}$$

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 $\Psi(w_1, \dots, w_M; z_1, \dots, z_N)$ 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).

25.3.2 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. 25.3 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 variable z , thus reducing the net angular momentum of each particle 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 lowest 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

$$\Psi_{qes}(w) = \left(\prod_{i=1}^N \left(2\ell^2 \frac{\partial}{\partial z_i} - w^* \right) \right) \phi \prod_{i=1}^N e^{-|z_i|^2/(4\ell^2)}$$

25.3.3 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 magnetic field incurs phases. We would like thus 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. 25.2



Fig. 25.2 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.

25.4 Digression on Berry's Phase

The Berry phase²⁰ is one of the most fundamental ideas of modern physics. We recall the adiabatic theorem. If you start in an eigenstate and change a Hamiltonian sufficiently slowly, and there are no level crossings, then the system will just track the eigenstate as it slowly changes — i.e., it remains in the instantaneous eigenstate. However, during this process it takes a bit of thought to figure out what happens to the phase 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

²⁰Berry's work on Berry Phase in 1984 had a number of precursors, most notably the work of Pancharatnam in 1956.

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 wavefunction 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

$$\gamma(t_f) - \gamma(t_i) = -\frac{1}{\hbar} \int_{t_i}^{t_f} E(\mathbf{R}(t)) dt - i \int_{\mathbf{R}_i}^{\mathbf{R}_f} d\mathbf{R} \cdot \langle \psi(\mathbf{R}) | \nabla_{\mathbf{R}} | \psi(\mathbf{R}) \rangle$$

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 it is a *geometric* phase.

25.5 Arovas-Schrieffer-Wilczek Calculation of Fractional Statistics

²¹Wilczek won a Nobel for his work on asymptotic freedom. Schrieffer won a Nobel for his work on BCS theory of superconductivity. Arovas was a grad student at the time.

This section follows the approach of Arovas, Schrieffer and Wilczek²¹.

Let us consider a $\nu = 1/m$ wavefunction for a quasihole

$$\Psi(w) = \mathcal{N}(|w|) \left[\prod_{i=1}^N (z_i - w) \right] \Psi_{Laughlin}^{(m)}$$

and we will imagine moving around the position w in a circle of constant radius as shown in the right of Fig. 25.2. 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 parameterize²² the position of the particle by the angle θ and $w = |w|e^{i\theta}$.

The Berry phase from moving the particle in a loop will then be

$$\Delta\gamma = -i \int_0^{2\pi} d\theta \langle \Psi(\theta) | \partial_\theta | \Psi(\theta) \rangle$$

where we have written $|\Psi(\theta)\rangle$ to mean $|\Psi(|w|e^{i\theta})\rangle$. We then have

$$\partial_\theta |\Psi(\theta)\rangle = \frac{\partial w}{\partial \theta} \left(\sum_i \frac{-1}{z_i - w} \right) |\Psi(\theta)\rangle$$

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 phase²³

$$\begin{aligned} \Delta\gamma &= -i \oint d\theta \langle \Psi(\theta) | \partial_\theta | \Psi(\theta) \rangle \\ &= -i \oint dw \sum_i \left\langle \Psi(w) \left| \frac{-1}{z_i - w} \right| \Psi(w) \right\rangle \end{aligned}$$

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. 25.2. Using a normalization factor that is again a function of $|w|$ only, the same calculation holds, but now the number of electrons enclosed has changed by one quasiparticle charge e/m . Thus the phase is now

$$\Delta\gamma = \gamma_{AB} + \gamma_{\text{statistical}}$$

where the additional phase for having gone around another quasihole is given by

$$\gamma_{\text{statistical}} = 2\pi/m$$

or in other words we have fractional statistics! For example, for the Laughlin state at $\nu = 1/2$, we have semionic statistics.

²³The 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! ***

²²On 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 ***

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 Aharonov-Bohm phase for taking a particle around flux, added to the statistical 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 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 situation 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.

25.6 Gauge Choice and Monodromy

The Laughlin wavefunction with M quasiholes takes the form

$$\Psi(w_1, \dots, w_M; z_1, \dots, z_N) = \mathcal{N}(w_1, \dots, w_N) \left[\prod_{\alpha=1}^M \prod_{i=1}^N (z_i - w_\alpha) \right] \Psi_{\text{Laughlin}}^{(m)}(z_1, \dots, z_N) \quad (25.4)$$

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

$$|\mathcal{N}(w_1, \dots, w_M)| = C \prod_{\alpha < \beta} |w_\alpha - w_\beta|^{1/m} \prod_{\alpha=1}^M e^{-|w_\alpha|^2/(4\ell^{*2})}$$

where C is some constant and

$$\ell^* = \sqrt{\frac{\hbar}{e^* B}}$$

is the effective magnetic length for a particle of charge $e^* = e/m$. This choice of normalization assures that

$$\langle \Psi(w_1, \dots, w_M) | \Psi(w_1, \dots, w_M) \rangle$$

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 interesting different convention is to choose

$$\mathcal{N}(w_1, \dots, w_N) = C \prod_{\alpha < \beta} (w_\alpha - w_\beta)^{1/m} \prod_{\alpha=1}^M e^{-|w_\alpha|^2/(4\ell^2)} \quad (25.5)$$

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.

25.6.1 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 position w and another at position w' the wavefunction is

$$\begin{aligned} \Psi(w) = & C(w - w')^{1/m} e^{-(|w|^2 + |w'|^2)/(4\ell^2)} \times \\ & \prod_i (z_i - w)(z_i - w') \prod_{i < j} (z_i - z_j) \prod_i e^{-|z_i|^2/(4\ell^2)} \end{aligned}$$

with C chosen so that Ψ is normalized independent of the quasihole coordinates.²⁴ Let us parameterize the path of a quasiparticle as $w(\tau)$.

²⁴Strictly 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.

We can write the Berry phase as

$$\Delta\gamma = -i \oint d\tau \langle \Psi(\tau) | \partial_\tau | \Psi(\tau) \rangle$$

We write

$$\frac{\partial}{\partial \tau} = \frac{\partial w}{\partial \tau} \frac{\partial}{\partial w} + \frac{\partial w^*}{\partial \tau} \frac{\partial}{\partial w^*} \quad (25.6)$$

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 $\partial/\partial w$ 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

$$\begin{aligned} \langle \Psi(w) | \partial_w | \Psi(w) \rangle &= \partial_w [\langle \Psi(w) | \Psi(w) \rangle] - [\partial_w \langle \Psi(w) |] | \Psi(w) \rangle \\ &= \frac{w^*}{4\ell^{*2}} \langle \Psi(w) | \Psi(w) \rangle = \frac{w^*}{4\ell^{*2}} \end{aligned}$$

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. 25.6. We now use the complex version of Stokes theorem²⁵ to obtain

²⁵The complex version of Stokes is as follows. Using $w = x + iy$

$$\begin{aligned} &\int_{\partial A} (F dw - G dw^*) \\ &= 2i \int_A (\partial_{w^*} F + \partial_w G) dx dy \end{aligned}$$

$$\Delta\gamma = \frac{\text{Area}}{\ell^{*2}} = 2\pi(1/m)\Phi/\phi_0$$

which is the Aharonov-Bohm phase corresponding to the flux enclosed in the path – without giving the fractional statistical phase which has now been moved to the monodromy!

The key point here, which we emphasize, is that if we work with normalized holomorphic wavefunctions (i.e., holomorphic gauge), then the fractional statistics are fully explicit in the monodromy of the wavefunction — we can read the statistics off from the wavefunction without doing any work!

25.7 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

$$j^\mu = \frac{-1}{2\pi m} \epsilon^{\mu\nu\lambda} \partial_\nu \delta A_\lambda \quad (25.7)$$

We must, of course have charge conservation as well. This is easy to enforce by writing the current in the form

$$j^\mu = \frac{1}{2\pi} \epsilon^{\mu\nu\lambda} \partial_\nu a_\lambda \quad (25.8)$$

which then automatically satisfies

$$\partial_\mu j^\mu = 0$$

In this language, the effective Lagrangian that produces Eq. 25.7 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 associated with the local gauge symmetry will be

$$j^\mu = \frac{\partial \mathcal{L}}{\partial A_\mu}$$

which matches the expression from Eq. 25.8. By differentiating the Lagrangian with respect to a_μ we generate the equations of motion Eq. 25.7.

More here

25.8 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 25.4, 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. 25.4. 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

$$\tilde{\Psi}(z_1, \dots, z_N) = \int \mathbf{d}w_1, \dots, \mathbf{d}w_M \phi^*(w_1, \dots, w_M) \Psi(w_1, \dots, w_M; z_1, \dots, z_N)$$

The general idea of this scheme is that the pseudo-wavefunction can itself be of the form of a Laughlin wavefunction. In the original Laughlin argument we wrote down wavefunctions for both boson and fermion particles. Here, the particles w are anyons, so we need to write a slightly different form of a wavefunction. We expect

$$\phi(w_1, \dots, w_M) = \prod_{\alpha < \beta} (w_\alpha - w_\beta)^{\frac{1}{m} + p}$$

with p an even integer. The fractional power accounts for the fact that the anyon wavefunction must be multi-valued as one particle moves around another. The factor p is to include a “Laughlin” factor repelling these anyons from each other without further changing the statistics.

The condensation of these quasi-particles into a Laughlin state generates a wavefunction for the filling fraction

$$\nu = \frac{1}{m \pm 1/p}$$

with the \pm corresponding to whether we are condensing quasiparticles or quasiholes. One can continue the argument starting with these new fractions and generating further daughter states and so forth. At the next level for example, we have

$$\nu = \frac{1}{m \pm \frac{1}{p \pm \frac{1}{q}}}$$

By repeating the procedure, any odd denominator fraction $\nu = p/q$ can be obtained.

Exercises

Exercise 25.1 Filled Lowest Landau Level

Show that the filled Lowest Landau level of non-interacting electrons (a single Slater determinant) can be written as

$$\Psi_m^0 = \mathcal{N} \prod_{1 \leq i < j \leq N} (z_i - z_j)^1 \prod_{1 \leq i \leq N} e^{-|z_i|^2/4\ell^2} \quad (25.9)$$

with \mathcal{N} some normalization constant. I.e., this is the Laughlin wavefunction with exponent $m = 1$.

Exercise 25.2 Laughlin Plasma Analogy

Consider the Laughlin wavefunction for N electrons at positions z_i

$$\Psi_m^0 = \mathcal{N} \prod_{1 \leq i < j \leq N} (z_i - z_j)^m \prod_{1 \leq i \leq N} e^{-|z_i|^2/4\ell^2} \quad (25.10)$$

with \mathcal{N} a normalization constant. The probability of finding particles at positions $\{z_1, \dots, z_N\}$ is given by $|\Psi_m^0(z_1, \dots, z_N)|^2$.

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

$$\beta v(\vec{r}_i - \vec{r}_j) = -2m \log(|\vec{r}_i - \vec{r}_j|) \quad (25.11)$$

in the presence of a background potential u such that

$$\beta u(|\vec{r}|) = |\vec{r}|^2/(2\ell^2) \quad (25.12)$$

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 $\{\vec{r}_1, \dots, \vec{r}_N\}$ is given by $|\Psi_m^0(z_1, \dots, z_N)|^2$ where $z_j = x_j + iy_j$ is the complex representation of position \vec{r}_j . 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, \dots, w_M .

$$\Psi_m = \mathcal{N}(w_1, \dots, w_M) \left[\prod_{1 \leq i \leq N} \prod_{1 \leq \alpha \leq M} (z_i - w_\alpha) \right] \Psi_m^0 \quad (25.13)$$

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

$$|\mathcal{N}(w_1, \dots, w_M)| = \mathcal{K} \prod_{1 \leq \alpha < \gamma \leq M} |w_\alpha - w_\gamma|^{1/m} \prod_{1 \leq \alpha \leq M} e^{-|w_\alpha|^2/(4m\ell^2)} \quad (25.14)$$

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).

Fractional Quantum Hall Edges

26.1 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 absence of the confinement.

Since the confinement is rotationally symmetric, we can still classify all eigenstates by their angular momentum quantum numbers. Using symmetric gauge we can still write the single particle eigenstates as¹

$$\varphi_m \sim z^m e^{-|z|^2/(4\ell^2)}$$

where m is the eigenvalue of the angular momentum² 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.

¹Note that the parabolic confinement modifies the magnetic length.

²We drop the \hbar from the angular momentum operator so its eigenvalues are just numbers.

26.2 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

$$\Psi_{Laughlin}^{(m)} = \prod_{i < j} (z_i - z_j)^m \prod_{i=1}^N e^{-|z_i|^2 / (4\ell^2)}$$

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 multiplying 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

$$\Psi = (\text{Any Symmetric Polynomial}) \Psi_{Laughlin}^{(m)}$$

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 enumerate 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, \dots, 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 functions³. 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 26.1.

Thus the number of edge excitations at a given angular momentum follows a pattern, 1, 2, 3, 5, 7, ... with energy increasing linearly with the

³In 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	$p_2, p_1 p_1$	2α
3	3	$p_3, p_2 p_1, p_1 p_1 p_1$	3α
4	5	$p_4, p_3 p_1, p_2 p_1 p_1, p_1 p_1 p_1 p_1$	4α
5	7	$p_5, p_4 p_1, p_3 p_2, p_3 p_1 p_1, p_2 p_2 p_1, p_2 p_1 p_1 p_1, p_1 p_1 p_1 p_1 p_1$	5α

Table 26.1 Table of Symmetric Polynomials

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)

26.2.1 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 elementary 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 26.2. 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 positive angular momentum $m > 0$ not negative angular momentum.⁵

⁴To 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

$$\dots \bullet \bullet \bullet \bullet \bullet \bullet \circ \circ \circ \circ \dots$$

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

$$\dots \bullet \bullet \bullet \bullet \bullet \bullet \circ \bullet \circ \circ \circ \dots$$

adding two units can be done in two ways

$$\dots \bullet \bullet \bullet \bullet \bullet \circ \bullet \bullet \circ \circ \circ \dots$$

and

$$\dots \bullet \bullet \bullet \bullet \circ \bullet \bullet \bullet \circ \circ \circ \dots$$

thus starting the series 1, 2, 3, 5, 7, ...

⁵An *achiral* bose field on a circle requires both positive and negative angular mo-

$L - L_{ground}$	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 26.2 Fock Space for Chiral Bosons

26.3 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.

mentum modes).

Conformal Field Theory

Approach to Fractional Quantum Hall Effect

27

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 intimate 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.

27.1 The Chiral Boson and The Laughlin State

An interesting feature of theories in 1+1d is that they can often be decomposed (mostly¹) cleanly into right moving and left moving pieces. So for example, if we take the simplest possible 1+1 d system, a free

¹There 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

$$H = \sum_{m>0} (\alpha m) a_m^\dagger a_m$$

²We have dropped the zero mode here.

Quantizing the boson lagrangian we find that²

$$\phi(x) = \sum_{m>0} \frac{i}{\sqrt{m}} e^{2\pi i m x / L} a_m^\dagger + \text{h.c.} \quad (27.1)$$

³Perhaps the easiest way to see this is to calculate directly from Eq. 27.1. 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

$$S = (8\pi)^{-1} \int dx dy |\nabla \Phi|^2$$

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.

⁴The 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.

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 correlator³

$$\langle \phi(z) \phi(z') \rangle = -\log(z - z')$$

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 ordering⁴ A straightforward exercise (assigned as homework!) using Wick's theorem then shows that

$$\begin{aligned} \langle V_{\alpha_1}(z_1) V_{\alpha_2}(z_2) \dots V_{\alpha_N}(z_N) \rangle &= e^{-\sum_{i<j} \alpha_i \alpha_j \langle \phi(z_i) \phi(z_j) \rangle} \\ &= \prod_{i<j} (z_i - z_j)^{\alpha_i \alpha_j} \end{aligned} \quad (27.2)$$

so long as

$$\sum_i \alpha_i = 0 \quad (27.3)$$

(otherwise the correlator vanishes).

27.1.1 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

$$\Psi_{Laughlin}^{(m)} = \langle \psi_e(z_1) \psi_e(z_2) \dots \psi_e(z_N) \hat{Q} \rangle = \prod_{i < j} (z_i - z_j)^m$$

The index α must be chosen such that α^2 is an integer such that the wavefunction is single valued in the electron coordinates. Note that here although the correlator means a 1+1d theory, we are constructing a wavefunction for a 2d system at fixed time!

Here, the operator \hat{Q} can be chosen in two different ways. One possibility is to choose $\hat{Q} = V_{-N\alpha}$, i.e., a neutralizing charge at infinity such that Eq. 27.3 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.

27.1.2 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

$$\begin{aligned} \Psi_{qh}(w) &= \langle \psi_{qh}(w) \psi_e(z_1) \psi_e(z_2) \dots \psi_e(z_N) \hat{Q} \rangle \\ &= \left[\prod_i (z_i - w)^{\beta\sqrt{m}} \right] \Psi_{Laughlin}^{(m)} \end{aligned} \quad (27.4)$$

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

$$e^{-|w|^2/(4m\ell^2)}$$

⁵To 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

$$\sum_{i, \beta} \log(|z_i - z_\beta|) \rightarrow \int d^2r \log(|z - r|)$$

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.

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

$$\begin{aligned} \Psi(w_1, \dots, w_M) &= \langle \psi_{qh}(w_1) \dots \psi_{qh}(w_M) \psi_e(z_1) \dots \psi_e(z_N) Q \rangle \quad (27.5) \\ &= C \prod_{\alpha < \beta} (w_\alpha - w_\beta)^{1/m} \prod_{\alpha=1}^M e^{-|w_\alpha|^2/(4\ell^*{}^2)} \left[\prod_{\alpha=1}^M \prod_{i=1}^N (z_i - w_\alpha) \right] \Psi_{Laughlin}^{(m)} \end{aligned}$$

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

$$V_{1/\sqrt{m}} \times V_{1/\sqrt{m}} \rightarrow V_{2/\sqrt{m}} \quad (27.6)$$

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 straightforwardly. 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 complicated TQFTs.

27.2 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 dimension⁶. They are generically

⁶We 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.

highly interacting theories, and most often it is impossible to write an explicit Lagrangian for the theory, but due to the special properties of being in 1+1 and having conformal invariance (guaranteed by being gapless in 1+1 d) these models are exactly solvable.

A particular CFT is defined by certain information known as conformal data, which basically mimics the defining features of a TQFT:

(1) There will be a finite set⁷ of so-called **primary fields**, which we might call $\phi_i(z)$ (or we may use other notation). These are analogous to the particle types in a TQFT. Every CFT has an identity field often called I (which isn't really a function of position). Correlators of these fields

$$\langle \phi_{j_1}(z_1) \dots \phi_{j_N}(z_N) \rangle$$

are always holomorphic functions of the z arguments, although there may be branch cuts.

(2) Each primary field has a **scaling dimension**⁸ or **conformal weight** or **conformal spin**, which we call h_i . The scaling dimension of I is $h_I = 0$. We have seen 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

$$I \times \phi_j = \phi_j$$

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

⁷A *nonrational* CFT may have an infinite number of particle types, but these are badly behaved and do not appear to correspond to nice TQFTs.

⁸In 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!

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

$$\lim_{w \rightarrow z} \phi_i(w) \phi_j(z) = \sum_k C_{ij}^k (w - z)^{h_k - h_i - h_j} \phi_k(z) + \dots$$

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 singular 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

$$\lim_{w \rightarrow z} \langle \psi_a(w) \psi_b(z) \psi_c(y_1) \psi_d(y_2) \dots \psi_n(y_m) \rangle = \sum_k C_{ab}^k (w - z)^{h_k - h_a - h_b} \langle \psi_k(z) \psi_c(y_1) \psi_d(y_2) \dots \psi_n(y_m) \rangle$$

27.2.1 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. 27.6. ***(clean this up)**

27.2.2 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 \dots 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 \dots 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 interesting OPE⁹

$$\sigma(w)\sigma(z) \sim C_{\sigma\sigma}^I (w-z)^{-1/8} I + C_{\sigma\sigma}^\psi (w-z)^{3/8} \psi(z) + \dots \quad (27.7)$$

where all terms in the \dots must have branch cuts that match one of the two leading terms.

Let us consider calculating a correlator,

$$\lim_{w \rightarrow z} \langle \sigma(w)\sigma(z) \rangle$$

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

$$\lim_{w \rightarrow z} \langle \sigma(w)\sigma(z) \rangle \sim (w-z)^{-1/8} \quad (27.8)$$

⁹Remember these exponents of 1/8 and 3/8 from the Ising anyon homework problems? ***

On the other hand, calculating

$$\lim_{w \rightarrow z} \langle \sigma(w) \sigma(z) \psi(y) \rangle$$

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

$$\lim_{w \rightarrow z} \langle \sigma(w) \sigma(z) \psi(y) \rangle \sim (w - z)^{3/8} \quad (27.9)$$

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. 27.8, whereas in the presence of any odd number of ψ fields it will be like Eq. 27.9.

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

¹⁰Insert footnote or appendix that derives this. See Yellow Book for now!

¹¹Add footnote on wick's theorem?***

27.3 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 multi-particle wavefunction

$$\Psi = \langle \psi_e(z_1)\psi_e(z_2)\dots\psi_e(z_N)Q \rangle$$

where Q is the background charge for the bose field. Since the Ising and bose fields are completely separate 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

$$\begin{aligned} \langle \psi(z_1)\psi(z_2)\dots\psi(z_N) \rangle &= \mathcal{A} \left[\frac{1}{z_1 - z_2} \frac{1}{z_3 - z_4} \dots \frac{1}{z_{N-1} - z_N} \right] \\ &\equiv \text{Pf} \left(\frac{1}{z_i - z_j} \right) \end{aligned} \quad (27.10)$$

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

$$\Psi = \text{Pf} \left(\frac{1}{z_i - z_j} \right) \prod_{i<j} (z_i - z_j)^m \prod_{i=1}^N e^{-|z_i|^2/(4\ell^2)}$$

which is known as the Moore-Read wavefunction. For m odd this is a wavefunction for 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 $\nu = 1/m$.

27.3.1 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

¹²Several interesting facts about the Pfaffian: A BCS wavefunction for a spinless superconductor can be written as $\text{Pf}[g(\mathbf{r}_i - \mathbf{r}_j)]$ where g is the wavefunction for a pair of particles. Any antisymmetric matrix M_{ij} has a Pfaffian

$$\text{Pf}[M] = \mathcal{A}[M_{12}M_{34}\dots].$$

Also it is useful to know that $(\text{Pf}[M])^2 = \det M$.

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 point¹³.

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 interaction

$$V = V_0 \sum_{i < j < k} \delta(\mathbf{r}_i - \mathbf{r}_j) \delta(\mathbf{r}_i - \mathbf{r}_k)$$

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 state¹⁴ 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.

27.4 Quasiholes of the Moore-Read state

We now try to construct quasiholes for the Moore-Read Pfaffian wavefunction. As we did in Eq. 27.4, 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

$$\psi_{qh}^L(w) = V_\beta(w)$$

¹³To see this, note that taking the first two particles to the same point gives

$$\lim_{z_2 \rightarrow z_1} \psi_e(z_1) \psi_e(z_2) \sim IV_2(z_1)$$

Then fusing the third particle

$$\lim_{z_3 \rightarrow z_1} \psi_e(z_3) V_2(z_1) \sim (z_3 - z_1)^2 \psi V_3(z_1)$$

¹⁴Here 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.

¹⁵There 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

Looking at the OPE we have (**include fields on the right? **)

$$\psi_{qh}^L(w)\psi_e(z) \sim (w-z)^{\beta\sqrt{m}}\psi(z)$$

In order to have the correlator be single valued in z (i.e., no branch cuts) we must choose $\beta = p/\sqrt{m}$ for some integer p (the smallest quasihole of this type corresponding to $p = 1$ then). This generates the wavefunction

$$\begin{aligned}\Psi_{qh}^L(w) &= \langle \psi_{qh}^L(w) \psi_e(z_1) \psi_e(z_2) \dots \psi_e(z_N) \hat{Q} \rangle \quad (27.11) \\ &= \left[\prod_{i=1}^N (z_i - w) \right] \Psi_{Moore-Read}^{(m)}\end{aligned}$$

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

$$\psi_{qh}(w)\psi_e(z) \sim [\sigma(w)\psi(z)] [V_\beta(w)V_{\sqrt{m}}(z)] \sim (w-z)^{-1/2}(w-z)^{\beta\sqrt{m}}$$

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 27.2, the operators must fuse to the identity in order to give a nonzero result. Thus, we must always have an even number of σ fields¹⁶. We thus consider the wavefunction of the form

$$\Psi_{qh}(w, w') = \langle \psi_{qh}(w)\psi_{qh}(w') \psi_e(z_1)\psi_e(z_2) \dots \psi_e(z_N) \hat{Q} \rangle \quad (27.12)$$

$$= (w-w')^{\frac{1}{4m}} e^{-(|w|^2+|w'|^2)/4\ell^2} \prod_{i=1}^N (w-z_i)^{1/2} (w'-z_i)^{1/2} \quad (27.13)$$

$$\times \langle \sigma(w)\sigma(w') \psi(z_1)\psi(z_2) \dots \psi(z_N) \rangle \prod_{i<j} (z_i - z_j)^m \prod_{i=1}^N e^{-|z_i|^2/(4\ell^2)}$$

Several comments are in order here. First of all, from the first line of Eq. 27.13 it looks like there are branch cuts with respect to the z coordinates. However, these fractional powers are precisely canceled by

¹⁶Like the Sith, they come in pairs.

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. 27.10). 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 coordinates (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 nontrivial¹⁷. 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!

¹⁷See work by Bonderson et al ***.

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. 27.8. As the two quasiholes approach each other we then have¹⁸ (** insert also h-h-h derivation of R? **)

¹⁸Strictly 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.

$$\psi_{qh}(w)\psi_{qh}(w') \sim (w - w')^{\frac{1}{4m} - \frac{1}{8}}$$

where the $\frac{1}{4m}$ is written explicitly in the first line of Eq. 27.13 and the $-\frac{1}{8}$ is from the operator product expansion Eq. 27.8. 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?)***

$$[R_{qh-qh}^{“I”}]^2 = e^{2\pi i(\frac{1}{4m} - \frac{1}{8})}$$

Recall that if $a \times b \rightarrow c$ we should have $[R_{ab}^c]^2 = e^{2\pi i(h_c - h_a - h_b)}$. 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 $h_{“I”} = 1/2m$.

Fusion and Braiding of Two Quasipoles 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 quasipoles fuse to ψ as in Eq. 27.9. As the two quasipoles approach each other we then have¹⁹

$$\psi_{qh}(w)\psi_{qh}(w') \sim (w - w')^{\frac{1}{4m} + \frac{3}{8}}$$

where the $\frac{1}{4m}$ is written explicitly in the first line of Eq. 27.13 and the $\frac{3}{8}$ is from the operator product expansion Eq. 27.9. Again we just read off the monodromy from this OPE. Thus, one obtains a different phase depending on the fusion channel of the two quasipoles. In the notation we defined in earlier chapters we have

$$[R_{qh-qh}^{“\psi”}]^2 = e^{2\pi i(\frac{1}{4m} + \frac{3}{8})}$$

¹⁹Strictly 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.

27.5 Multiple Fusion Channels and Conformal Blocks

We will next address the issue of what happens when we have more than two quasipoles. It is clear what will happen here, we will obtain a correlator (like that in Eq. 27.13) 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

$$G(w_1, w_2, w_3, w_4) = \langle \sigma(w_1)\sigma(w_2)\sigma(w_3)\sigma(w_4) \rangle \quad (27.14)$$

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 27.2). 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 explicit form of the correlator in Eq. 27.14. 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 possible²⁰.

²⁰In 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!

$$\lim_{w \rightarrow \infty} \langle \sigma(0)\sigma(z)\sigma(1)\sigma(w) \rangle = a_+ G_+(z) + a_- G_-(z) \quad (27.15)$$

where

$$G_{\pm} = (wz(1-z))^{-1/8} \sqrt{1 \pm \sqrt{1-z}} \quad (27.16)$$

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 $z \rightarrow 0$ we find that

$$\begin{aligned}\lim_{z \rightarrow 0} G_+ &\sim z^{-1/8} & (\sigma(0)\sigma(z) \rightarrow I) \\ \lim_{z \rightarrow 0} G_- &\sim z^{3/8} & (\sigma(0)\sigma(z) \rightarrow \psi)\end{aligned}$$

Thus (comparing to Eqs. 27.8 and 27.9) 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. 27.16 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 anyone!)

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 assumption²¹) 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.

²¹As 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. *** maybe 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 27.16 it naively looks like both

$$\begin{aligned} \lim_{z \rightarrow 1} G_+ &\sim (1-z)^{-1/8} \\ \lim_{z \rightarrow 1} G_- &\sim (1-z)^{-1/8} \end{aligned}$$

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

$$\lim_{z \rightarrow 1} \tilde{G}_+ \sim (1-z)^{-1/8}$$

$$\begin{aligned}\lim_{z \rightarrow 1} \tilde{G}_- &\sim (1-z)^{-1/8} \left[\sqrt{1 + \sqrt{1-z}} - \sqrt{1 - \sqrt{1-z}} \right] \\ &\sim (1-z)^{-1/8} (1-z)^{1/2} \sim (1-z)^{3/8}\end{aligned}$$

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. 27.1

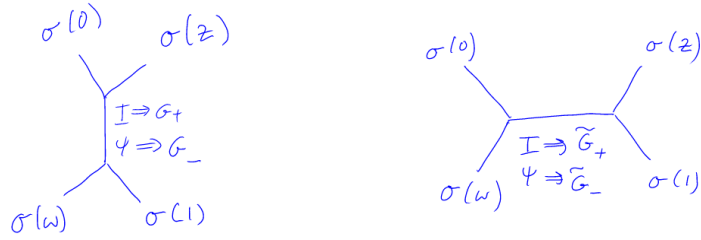


Fig. 27.1 The F -matrix transforms between the two fusion channels depicted here.

27.6 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. 27.13 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.

27.7 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 by starting 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.

27.7.1 \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	\times	ψ_1	ψ_2	σ_1	σ_2	ϵ	
ψ_1	$2/3$	ψ_1	ψ_2					
ψ_2	$2/3$	ψ_2	I	ψ_1				
σ_1	$1/15$	σ_1	ϵ	σ_2	$\sigma_2 + \psi_1$			
σ_2	$1/15$	σ_2	σ_1	ϵ	$I + \epsilon$	$\sigma_1 + \psi_2$		
ϵ	$2/5$	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

$$\begin{aligned}\sigma_1 &= \psi_2 \tau \\ \sigma_2 &= \psi_1 \tau\end{aligned}$$

$$\epsilon = \tau$$

²²Note 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 $h = 2/5 + 2/3 = 1 + 1/15$.

and using the Fibonacci fusions $\tau \times \tau = I + \tau$ and the \mathbb{Z}_3 fusions $\psi_i \times \psi_j = \psi_{(i+j) \bmod 3}$ with ψ_0 being the identity, we recover the full fusion table²². Let us propose an electron field

$$\psi_e(z) = \psi_1(z) V_{\sqrt{m + \frac{2}{3}}}(z)$$

where m is a nonnegative integer (even for bosons, odd for fermions). It is easy to check from the OPE that

$$\psi_e(z) \psi_e(w) \sim (z - w)^m \psi_2(z) V_{2\sqrt{m + \frac{2}{3}}}(z)$$

The resulting wavefunction is then

$$\Psi = \langle \psi_e(z_1) \psi_e(z_2) \dots \psi_e(z_N) Q \rangle$$

which is known as the Read-Rezayi \mathbb{Z}_3 parafermion wavefunction.

The filling fraction of the wavefunction is determined by the vertex operator and is given by

$$\nu = \frac{1}{m + \frac{2}{3}}$$

For the $m = 0$ case this is $\nu = 3/2$ bosons, while for the $m = 1$ case this is $\nu = 3/5$ fermions.

For the case of $m = 0$ it is easy to check that the wavefunction does not vanish when two particles come to the same point, nor does it vanish when three particles come to the same point, but it does vanish when four particles come to the same point. Thus the wavefunction is an exact (densest) zero energy ground state of a *four* particle delta function.

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 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 27.1 Bose Vertex Operators

In lecture we needed the following identity

$$\langle V_{\alpha_1}(z_1) V_{\alpha_2}(z_2) \dots V_{\alpha_N}(z_N) \rangle = \prod_{i < j} (z_i - z_j)^{\alpha_i \alpha_j} \quad (27.17)$$

where

$$\sum_i \alpha_i = 0 \quad (27.18)$$

where the vertex operators are defined by

$$V_\alpha(z) =: e^{i\alpha\phi(z)} : \quad (27.19)$$

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

$$e^{\alpha a} e^{\beta a^\dagger} = e^{\beta a^\dagger} e^{\alpha a} e^{\alpha\beta} \quad (27.20)$$

(b) Thus derive

$$\langle V_{A_1} V_{A_2} \dots V_{A_N} \rangle = e^{\sum_{i < j} \langle A_i A_j \rangle} \quad (27.21)$$

where

$$A_i = u_i a^\dagger + v_i a \quad (27.22)$$

and

$$V_{A_i} =: e^{A_i} := e^{u_i a^\dagger} e^{v_i a} \quad (27.23)$$

with the colons meaning normal ordering (all daggers moved to the left).

(c) Show that Eq. 27.21 remains true for any operators A_i that are sums of different bose modes a_k , i.e., if

$$A_i = \sum_k [u_i(k) a_k^\dagger + v_i(k) a_k] \quad (27.24)$$

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

$$\langle \phi(z) \phi(w) \rangle = -\ln(z - w) \quad (27.25)$$

conclude that Eq. 27.17 holds.

Note: This result is not quite correct, as it fails to find the constraint Eq. 27.18 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. Tsvetlik'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

$$\mathcal{L} = \frac{1}{2\pi} \partial_x \phi (\partial_x + v \partial_t) \phi \quad (27.26)$$

which clearly must be invariant under the global transformation $\phi \rightarrow \phi + b$.

(d) Show that the correlator Eq. 27.17 (with Eq. 27.19) cannot be invariant under this transformation unless Eq. 27.18 is satisfied, or unless the value of the correlator is zero.

Exercise 27.2 \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

$$(a \times b) \times c = a \times (b \times c) \quad (27.27)$$

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	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

$$\lim_{z' \rightarrow z} \psi_1(z') \psi_1(z) \sim (z' - z)^\alpha \psi_2(z) \quad (27.28)$$

(b) Construct all possible “electron” fields by making a product of the ψ_1 field and a chiral bose vertex operator of the form

$$\psi_e(z) = \psi_1(z) e^{i\beta\phi(z)} \quad (27.29)$$

that give a single-valued and nonsingular wavefunction for the electron. (See Eq. 27.17, but ignore the sum condition Eq. 27.18) 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

$$\phi_{qh}(w) = \varphi(w) e^{i\kappa\phi(w)} \quad (27.30)$$

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 quasipoles 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 quasipoles 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

$$\text{Degeneracy} \sim d^{[\text{Number of Quasiholes}]} \quad (27.31)$$

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

$$\psi_e(z) = \psi_2(z) e^{i\beta\phi(z)} \quad (27.32)$$

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).

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.

28.1 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, \dots, x_n) . Sometimes people call these spaces “Euclidean” space.

Definition 28.1 *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*.

28.1.1 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 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} .

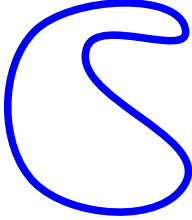


Fig. 28.1 This object is topologically a circle, S^1 .

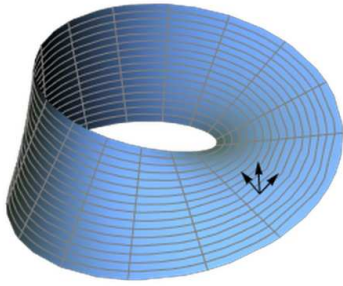


Fig. 28.2 A Möbius strip is a nonorientable manifold (with boundary). If we move the coordinate axes around the strip, when they come back to the same position, the normal vector will be pointing downwards instead of upwards.

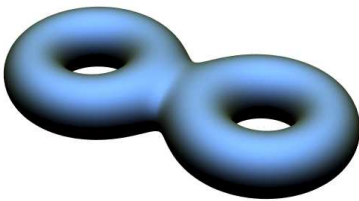


Fig. 28.3 A two handled torus is an orientable two-dimensional manifold without boundary. Because it has two holes we say it has genus two. Two dimensional manifolds without boundary are classified by their genus.

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. 28.1, 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. 28.2. If we smoothly 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. 28.3.

28.1.2 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.

28.1.3 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 doughnut¹ 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.

¹Alternatively spelled “donut” if you are from the states and you like coffee.

28.1.4 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 ends².

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 .

²In the notation of Section 28.1.5 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 28.1.2.

- 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.

28.1.5 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 $\partial\mathcal{M}$ 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

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

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.

28.2 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.,

Norwegian
died such
The word
italized due
e are a few
which are not
named after

if $g_1g_2 = g_2g_1$ 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 $g_1g_2 \neq g_2g_1$ then the group is called **nonabelian**⁴.

A group must always be associative

$$g_1(g_2g_3) = (g_1g_2)g_3 = g_1g_2g_3$$

Within the group there must exist an **identity** element which is sometimes⁵ called e or I or 0 or 1 . The identity element satisfies

$$ge = eg = g$$

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

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

⁵It 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.

28.2.1 Some Examples of Groups

- The group of integers \mathbb{Z} with the operation being addition. The identity element is 0. This group is abelian.
- The group $\{1, -1\}$ with the operation being the usual multiplication. This is also called the group \mathbb{Z}_2 . The identity element is 1. We could have also written this group as $\{0, 1\}$ with the operation being the usual addition modulo 2, where here the identity is 0. This group is abelian.
- The group \mathbb{Z}_N which is the set of complex numbers $e^{2\pi ip/N}$ with p an integer (which can be chosen between 1 and N inclusive) and the operation being multiplication. This is equivalent to the set of integers modulo N with the operation being addition. This group is abelian.
- The group of permutations of N elements, which we write as S_N (known as the **permutation group**, or **symmetric group**). This group is 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

$$X = \begin{cases} 1 & \rightarrow & 2 \\ 2 & \rightarrow & 1 \\ 3 & \rightarrow & 3 \end{cases}$$

Another element is

$$R = \begin{cases} 1 & \rightarrow & 2 \\ 2 & \rightarrow & 3 \\ 3 & \rightarrow & 1 \end{cases}$$

where X stands for exchange (exchanges 1 and 2) and R stands 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

$$XR = \begin{cases} 1 & \rightarrow & 1 \\ 2 & \rightarrow & 3 \\ 3 & \rightarrow & 2 \end{cases}$$

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

$$RX = \begin{cases} 1 & \rightarrow & 3 \\ 2 & \rightarrow & 2 \\ 3 & \rightarrow & 1 \end{cases}$$

It is easy to check that

$$X^2 = R^3 = e \quad (28.1)$$

and further we have

$$XR = R^2X \quad (28.2)$$

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 possibilities using Eqs. 28.1 and 28.2.

28.2.2 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 .

28.2.3 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:

- 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 invertible.
- 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 invertible) 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.

⁶Pronounced “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.

⁷A slightly more precise definition is that a Lie algebra is a vector space of elements u, v, w, \dots with coefficients in \mathbb{R} or \mathbb{C} such that $au + bv + cw + \dots$ is also in the space for all a, b, c, \dots . The bracket or commutator $[\cdot, \cdot]$ is a bilinear map from the space to itself such that $[X, aY + bZ] = a[X, Y] + b[X, Z]$ for all X, Y, Z and $[X, Y] = -[Y, X]$. The Jacobi identity $[[X, Y], Z] + [[Y, Z], X] + [[Z, X], Y] = 0$ must also hold.

A **Lie Algebra** is the algebra generated by elements infinitesimally close to the identity in a Lie group⁷. 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?

28.2.4 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

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

preserving multiplication. I.e.,

$$\rho(g_1)\rho(g_2) = \rho(g_1g_2)$$

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.

Irreducible representations matrices satisfy a beautiful orthogonality relationship known as the grand orthogonality theorem (or Schur

orthogonality)

$$\frac{1}{|G|} \sum_{g \in G} [\rho^R(g)]_{nm}^* [\rho^{R'}(g)]_{pq} = \delta_{np} \delta_{mq} \delta_{RR'} / d \quad (28.3)$$

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.

$$\chi_R(g) = \text{Tr}[\rho^R(g)]$$

where the superscript R indicates we are considering a particular representation R . Because of the cyclic property of the trace $\text{Tr}[ab] = \text{Tr}[ba]$ 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!

28.3 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.

28.3.1 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 $S^1 \times S^1$ 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 fundamental group is the group with two elements $\Pi_1(SO(3)) = \mathbb{Z}_2$.

⁸This is the origin of half-odd integer angular momenta.

⁹This 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].

28.4 Isotopy, Reidemeister Moves

We ran into the idea of isotopy in chapter 2. Two knots (or two pictures of knots) are isotopic if one can be deformed into each other other without cutting any of the strands. Usually this is referred to as “ambient isotopy”. In order for two pictures of knots to be ambient isotopic they must be related to each other by a series of moves, known as Reidemeister moves⁹.

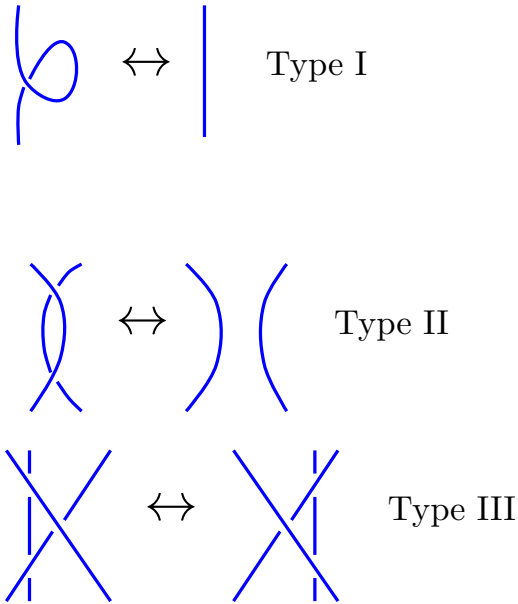
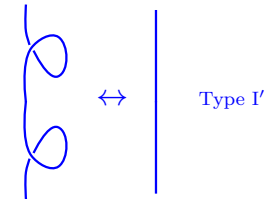


Fig. 28.4 The Three Reidemeister Moves. Any two knots that can be deformed into each other without cutting (they are “ambient isotopic”) can be connected by a series of Reidemeister moves.

It is also useful to define *regular isotopy* which is when two knots can be related to each other using only type-II and type-III moves. Another way of thinking about this is to think of the strings as being ribbons. A type-I move inserts a twist in the ribbon (See Fig. 2.7) and gives back

sion of the type I move that cancels two twists in opposite direction, as shown in Fig. 28.5



28.5 Linking and Writhe 331

Fig. 28.5 A Type I' move. This is an additional move needed for ambient isotopy of links.

a different ribbon diagram, whereas type-II and III moves do not twist the ribbon¹⁰.

28.5 Linking and Writhe

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 as shown in Fig. 28.6



Fig. 28.6 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 values of the crossings

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

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

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

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

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 Reidemeister moves, writhe, and linking, see

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

Exercises

Exercise 28.1 Reidemeister moves and the Kauffman Bracket Invariant

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

Exercise 28.2 Jones polynomial

Let us define the Jones polynomial of an oriented knot as

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

where w is the writhe. Show that this quantity is an invariant of ambient isotopy – that is, it is invariant under all three Reidemeister moves.

Commentary on References

- (1) A general reference which should be useful for much of the book is the review article by Nayak, Simon, 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 Kauffman [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. 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 Aharonov-Bohm effect. Look mainly at the first 22 pages.
- (6) The classic paper by Ed Witten which launched the field is ?. 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 mathematical viewpoint several articles by Sir Michael Atiyah are very useful. These are both introductions to topological quantum field theories. There is also a more detailed book by the same author. The full book might be hard to read unless you have a very strong maths background.
- (8) A rather remarkably well written and readable master's thesis (!) by Lokman Tsui on Chern-Simons theories, topological quantum field theory, and knot theory.
- (9) There are several nice references on the structure of topological quantum field theories and diagrammatic calculus,

Parsa Bonderson's thesis: <http://thesis.library.caltech.edu/2447/2/thesis.pdf>

This is a more detailed version of the long article by Kitaev ("Anyons in exactly solvable models") which I mention below. Note there is some slight change of convention between the two articles.

Also a good reference is the book on Topological Quantum Computation by Zhenghan Wang

"Topological Quantum Computation", Conference Board of the Mathematical Sciences, Regional Conference Series in Mathematics, American Mathematical Society, (Providence, Rhode Island), Number 112, 2008.

If you are more mathematical, you might like the thesis of Bruce Bartlett available online here

<https://arxiv.org/abs/math/0512103>

- (10) The monumental work "Anyons in an exactly solved model and beyond" by Alexei Kitaev, *Annals of Physics* 321 (2006) 2–111 available online here

<https://arxiv.org/abs/cond-mat/0506438>

This brings the ideas of topological quantum field theory into the condensed matter arena. This is not easy reading, but a ton of great ideas are buried in this paper.

Another work by Kitaev, "Fault-tolerant quantum computation by anyons", *Annals Phys.* 303 (2003) 2–30.

available online here

<https://arxiv.org/abs/quant-ph/9707021>

introduces the famous toric code, discusses quantum error correction, and generalizes the toric code model to arbitrary non-abelian groups.

Kitaev's work on the quantum wire (which we might get to at the end of the course) is here.

<https://arxiv.org/abs/cond-mat/0010440>

A brief digest of some of the many ideas introduced in these three papers is given by notes taken by Laumann of Kitaev's lectures, available here.

<https://arxiv.org/abs/0904.2771>

Loop gases are introduced in this paper by Freedman et al. It has a lot of sections which are hard to parse.

<http://stationq.cnsi.ucsb.edu/freedman/Publications/83.pdf>

The double-fibonacci string-net is discussed in some detail in this work by Fidkowski et al,

<https://arxiv.org/abs/cond-mat/0610583>

The classic paper on string - nets very generally is this by Levin and Wen.

<https://arxiv.org/abs/cond-mat/0404617>

The standard reference on introductory quantum hall effect is the classic book, "The Quantum Hall Effect", edited by Prange and Girvin, published by Springer. The first chapter, and the chapters by Laughlin and Haldane are probably the best. The experimental chapters are good for context too.

Another decent reference quantum Hall physics is T. Chakraborty and P. Piettiläinen, "The Quantum Hall Effects: Integral and Fractional," (Springer 1995).

A short review article by Macdonald is pretty nice and is available here.

<https://arxiv.org/pdf/cond-mat/9410047v1.pdf>

The article that introduced the ideas of conformal field theory into the field of quantum Hall effect is by Moore and Read, available online here.

<http://www.physics.rutgers.edu/gmoore/MooreReadNonabelions.pdf>

A recent review article on Fractional quantum Hall hierarchies (and also discusses nonabelian quantum Hall and conformal field theory) is online here.

<https://arxiv.org/abs/1601.01697>

A few random digressions:

- (11) If you are interested in 2+1 D quantum gravity, see this article .
I can't vouch for it, but the introduction is interesting;
<https://link.springer.com/article/10.12942/lrr-2005-1>
This is the article by Witten explaining how 2+1 D gravity is "exactly solvable." More from Witten here. There is reconsideration many years later, again by Witten, see here .
<http://www.sciencedirect.com/science/article/pii/0550321389905919>
- (12) I've been told the book by Jiannis Pachos on topological quantum computation is a good resource.
- (13) If you are interested in the topology of manifolds in 3 and 4 dimensions, there are several good books. One by Kirby is online here.
<https://math.berkeley.edu/kirby/papers/Kirby>
There is a book by Gompf and Stipsitz "4-manifolds and Kirby Calculus" which is nice. Note that parts of this book are online free if you google them.
<https://www.amazon.co.uk/4-Manifolds-Calculus-Graduate-Studies-Mathematics/dp/0821809946>
- (14) For more information on conformal field theory. The standard reference is the Big yellow book (Conformal Field Theory Authors: Philippe Di Francesco, Pierre Mathieu, David Senechal) . 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](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](https://arxiv.org/abs/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](https://doi.org/10.1103/PhysRev.115.485). URL <https://link.aps.org/doi/10.1103/PhysRev.115.485>.
- 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. *Publications Mathématiques de l’Institut des Hautes Scientifiques*, 68, 1988. doi: <https://doi.org/10.1007/BF02698547>.
- 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](https://doi.org/10.1016/0550-3213(80)90474-5).
- B. Bakalov and A. Kirillov. *Lectures on Tensor Categories and Modular Functors*, volume 21 of *University Lecture Series*. American Mathematical Society, 2001.
- 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>.
- P. H. Bonderson. *Non-Abelian Anyons and Interferometry*. PhD thesis, California Institute of Technology, 2007.
- 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](https://doi.org/10.1103/PhysRevB.80.155136). URL <https://link.aps.org/doi/10.1103/PhysRevB.80.155136>.

- 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. Crane and D. Yetter. A categorical construction of 4-d topological quantum field theories? In L. H. Kauffman and R. Baadhio, editors, *Quantum Topology*. 1993.
- 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>.
- P. Etingof, S. Gelaki, D. Nikshych, and V. Ostrik. *Tensor Categories*. American Mathematical Society, 2015.
- R. P. Feynman and A. R. Hibbs. *Quantum Mechanics and Path Integrals*. McGraw Hill, 1965. Reprinted 2005, Dover.
- 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.
- P. Freyd, D. Yetter, J. Hoste, W. B. R. Lickorish, K. Millett, and A. Ocneanu. *Il. 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.
- 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>.
- B. I. Halperin. Statistics of quasiparticles and the hierarchy of fractional quantized Hall states. *Phys. Rev. Lett.*, 52(18):1583–6, 1984.
- 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](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.
- 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.
- A. Y. Kitaev. Anyons in an exactly solved model and beyond. *Ann. Phys. (N.Y.)*, 321:2–111, 2006. cond-mat/0506438.
- 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 Mathematics*, 182:491–564, 2015. doi: <https://doi.org/10.4007/annals.2015.182.2.3>.
- J. M. Leinaas and J. Myrheim. On the theory of identical particles. *Nuovo Cimento*, 37B:1, 1977.
- 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.
- G. Moore and N. Read. Nonabelions in the fractional quantum Hall effect. *Nucl. Phys. B*, 360(2-3):362–96, 1991.

- 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>.
- H. Nicolai and K. Peeters. *Loop and Spin Foam Quantum Gravity: A Brief Guide for Beginners*, pages 151–184. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. 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>.
- 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>.
- 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.
- 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>.
- N. Y. Reshetikhin and V. G. Turaev. Invariants of 3-manifolds via link polynomials and quantum groups. *Invent. Math.*, 103:547?597, 1991.
- 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>.
- S. H. Simon. Quantum computing with a twist”. *Physics World*, pages 35–40, September 2010.
- 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. *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):865902, 1992.
- S. Vandoren and P. van Nieuwenhuizen. Lectures on instantons. *arxiv/0802.1862*, 2008.
- 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. *arXiv:1910.13441*, 2020.
- 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](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.