LINKS, QUANTUM GROUPS AND TQFTS

STEPHEN SAWIN

ABSTRACT. The Jones polynomial and the Kauffman bracket are constructed, and their relation with knot and link theory is described. The quantum groups and tangle functor frameworks for understanding these invariants and their descendents are given. The quantum group $U_q(sl_2)$, which gives rise to the Jones polynomial, is constructed explicitly. The 3-manifold invariants and the axiomatic topological quantum field theories which arise from these link invariants at certain values of the parameter are constructed and proven to be invariant.

INTRODUCTION

In studying a class of mathematical objects, such as knots, one usually begins by developing ideas and machinery to understand specific features of them. When invariants are discovered, they arise out of such understanding and generally give information about those features. Thus they are often immediately useful for proving theorems, even if they are difficult to compute.

This is not the situation we find ourselves in with the invariants of knots and 3-manifolds which have appeared since the Jones polynomial. We have a wealth of invariants, all readily computable, but standing decidedly outside the traditions of knot and 3-manifold theory. They lack a geometric interpretation and consequently have been of almost no use in answering questions one might have asked before their creation. In effect, we are left looking for the branch of mathematics from which these should have come organically.

In fact, we have an answer of sorts. In [Wit89a], Witten gave a heuristic definition of the Jones polynomial in terms of a topological quantum field theory, following the outline of a program proposed by Atiyah [Ati88]. The sketch of his construction in the following paragraph is fairly technical but is independent of the rest of the paper and can safely be skimmed for the flavor.

Witten considered a knot in a 3-manifold and a connection $A$ on some principal $G$-bundle, with $G$ a simple Lie group. The Chern-Simons functional associates a number $CS(A)$ to $A$, but it is gauge-invariant only up to an integer, so the quantity $\exp(2\pi i k CS(A))$ is gauge-invariant. Also, the holonomy of the connection around the knot is an element of $G$ well-defined up to conjugation, so the trace of it with respect to a given representation is well-defined. Multiplying these two gives a number depending on the knot, the manifold, the representation and the connection.
The magic comes when we average over all connections and all principal bundles: Of course, this makes no sense, since there is no apparent measure on the infinite-dimensional space of connections. But proceeding heuristically, such an average should depend only on the manifold, representation and the isotopy type of the knot. Witten argued, using a close correspondence with conformal field theory, that when the manifold was $S^3$, this invariant had combinatorial properties that forced it to be the analogue of the Jones polynomial for the given group and representation (specifically, the Jones polynomial at certain values of the parameter). Needless to say, a long physics tradition of very successful heuristic reasoning along these lines suggested to Witten that this ill-defined average should make sense in this case.

The fundamental problem of the field, then, is to develop this nonrigorous but beautiful geometric interpretation of the new invariants into a rigorous one that reasonably captures its flavor. This is important for at least two reasons.

First, such an interpretation of the invariants is probably necessary for their application to topological questions. While the invariants have yet to prove themselves in this regard, it is to be hoped that as part of a well-developed theory they could go a long way towards unlocking the mysteries of low-dimensional topology. This theory seems most likely to arise out of something like Witten’s construction, which relates the invariants explicitly to gauge theory and other geometry.

Second, Chern-Simons field theory is a field theory which is clearly nontrivial but which admits an exact combinatorial solution. In this sense, interpreting Witten’s work rigorously is just part of a large endeavor within mathematical physics to understand mathematically the whole of quantum field theory. Such an understanding seems far more tractable for Chern-Simons theory, with its combinatorial, finite-dimensional expression, than for genuinely physical theories. From this point of view topology may be seen as a laboratory for a particularly simple kind of physics, the understanding of which may point us towards a rigorous foundation for the more complex and physically interesting theories.

Work towards this goal proceeds along two fronts. On the one hand are efforts to understand the geometry and physics of Chern-Simons field theory, most notably recent efforts to do perturbation theory in this context by Axelrod-Singer and Kontsevich [AS92, AS94, Kon94]. On the other are combinatorial, algebraic, and topological efforts to understand the invariants themselves, especially with an eye towards structures suggested by the physics. This is the subject of the present article.

We begin in Section 1 with a discussion of the Jones polynomial, or more precisely the Kauffman bracket form of it. We mention other invariants, but throughout our focus will be on this example as the easiest. Section 2 first gives a functorial framework for viewing these invariants, constructing the functor explicitly for the Kauffman bracket, and then shows such a functorial setup arises naturally from a certain algebraic structure: a ribbon Hopf algebra. Section 3 discusses quantum groups, the family of ribbon Hopf algebras associated to each Lie algebra. We construct the quantum group associated to $sl_2$ explicitly and show that with the fundamental representation it gives the Kauffman bracket as its invariant.

The last four sections sketch the construction of the combinatorial version of Chern-Simons theory. These bear further comment. Witten’s construction obviously gives a 3-manifold invariant as well, by considering an arbitrary manifold with the empty link. This formed much of the inspiration for Reshetikhin and Turaev’s construction of the 3-manifold invariants [RT91] using the algebraic machinery of
quantum groups. But Atiyah observed that this averaging process (called path integration by physicists), if taken at its word, implied some strong statements about how the invariants behave under cutting and pasting. These statements were formulated by Atiyah into his axioms for a topological quantum field theory, or TQFT [Ati89, Ati90a]. It was shown by Walker [Wal] and Turaev [Tur94] that Reshetikhin and Turaev’s 3-manifold invariants satisfy these axioms, although the key ideas of this proof already appeared in [RT91] and in a less rigorous form in Witten [Wit89b, Wit89a]. One of the principal aims of this paper is to sketch a simple proof of this fact, which is the best rigorous connection we have between these invariants and actual quantum field theory.

Section 4 discusses the properties of the quantum groups at roots of unity, summarized as their being modular Hopf algebras, which allow one to construct TQFTs from them. Section 5 gives a categorical formulation of TQFTs. Section 6 sketches a purely combinatorial description of the relevant category, that of biframed 3-dimensional cobordisms, in terms of surgery on links. This allows us in Section 7 to construct the TQFT out of the link invariants we have already defined.

Each section begins with a few paragraphs summarizing the results of the section as well as their motivation and larger context.

This paper is intended to introduce and invite a large mathematical audience to this exciting field. The focus is on giving the flavor and illustrating some of the power of a few simple ideas. This is attempted by proving an important result with a minimum of machinery, rather than by surveying the whole of the field or giving a full account of the machinery. Consequently many interesting areas go unmentioned or barely mentioned. By the same token, the references are chosen primarily to point the interested reader to good introductory accounts, rather than to give an exhaustive assignment of credit. I apologize for any omissions and point the reader to those same introductory works, many of which have excellent bibliographies, for the complete story.

I would like to thank Scott Axelrod, John Baez, Joan Birman, Dana Fine, Vaughan Jones, Louis Kauffman, Robert Kotiuga, Haynes Miller, Andrew Pressley, Justin Roberts, Lisa Sawin, Isadore Singer, James Stasheff, Vladimir Turaev, David Vogan and Eric Weinstein for helpful conversations, comments and suggestions.

1. Links and their invariants

The study of knots and links begins with some physics which, while a little eccentric sounding to modern ears and decidedly wrong, bears a striking resemblance to the physics where this story ends. In 1867 Lord Kelvin proposed that atoms were knotted vortices of ether and molecules were linked atoms [Kel67]. Work began to solve the basic problem of knot theory: When are two knots the same? Early efforts to approach this by careful tabulating and naive searching for invariants failed completely. Modern topology has proven more successful. The strategy of focusing exclusively on the knot complement and almost exclusively on its fundamental group has, by dint of great effort, produced a complete but thoroughly impractical algorithm for determining if two prime knots (knots which cannot be cut into two smaller knots) are the same up to orientation of the knot and space [Hem92]. The story for links is more complicated and not as well understood.

The Jones polynomial, with its descendants, grew from an entirely different branch of mathematics. Its origins bear telling as a study in serendipity, although
we will only sketch Jones’ construction, which is buried in the construction in Section 2. Jones, in the course of proving an important result about the ways in which certain algebras of operators sit inside each other [Jon83], constructed an algebra with a trace on it. He then noticed that this algebra gave a representation of the braid group (the mapping-class group of the plane with finitely many points removed). Now by Markov’s theorem [BZ85], links can be represented by elements of the braid group, with two braids giving the same link exactly when they can be connected by conjugation and another move, called stabilization. Any trace is invariant under conjugation, and this particular trace could easily be normalized so as to be invariant under stabilization. Thus presenting the link as a braid, representing the braid in the algebra, and taking the trace together give an invariant of the link. Since the algebra and the trace depend on a parameter $t$, so does the invariant, and it turns out the invariant is essentially a Laurent polynomial in $t$: the celebrated Jones polynomial.

This invariant almost certainly does not come from classical knot theory: It distinguishes links with diffeomorphic complement and detects mirror images for example. It is eminently computable: Although the algorithm is roughly exponential in the number of crossings, and the problem is known to be $\#P$-hard [JVW90], a clever high school student can compute it easily. The collection of these invariants is a fairly good distinguisher of knots, though it does not distinguish all knots, and it is not even known whether it can distinguish a knot from the unknot [Gar]. Maddeningly, these invariants have no geometric interpretation (at least, no rigorous one), and consequently they have provided few purely topological results (chiefly the Tait conjectures, about so-called alternating knots). With such mysteries before us, it is time to do some mathematics.

A link is a smooth embedding of several copies of $S^1$ into oriented $S^3$. A knot is a link with one component. Links are equivalent if there is a smooth isotopy of $S^3$ taking one to the other. This notion is equivalent to the corresponding PL notion and to the corresponding continuous notion if we restrict to ‘tame’ knots [BZ85]. An oriented link has an orientation on each component, and a framed link comes equipped with a nonzero section of the normal bundle, also up to smooth isotopy. A framed link may be thought of as formed from thick cord, so that cutting the cord, twisting it, and gluing it back together changes the framed link even though it does not change it as a link. We will draw links as in Figure 1, with only transverse-double-point crossings (such drawings are called projections of the link). Oriented links will be drawn with arrows indicating orientation, and framed links will be drawn so that the framing lies in the plane (thus in Figure 1, B is A with a full twist added). The unknot is the knot which bounds a disk and has an obvious preferred framing (shown in A). This is all the knot theory we will need to define the Jones polynomial, using a construction due to Kauffman [Kau87].

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{some-knots-and-links.png}
\caption{Some knots and links}
\end{figure}
Let $L$ be a projection of a framed, unoriented link $L$. The Kauffman bracket of $L$, $\langle L \rangle$, is an element of $\mathbb{Z}[A, A^{-1}]$, with $A$ an indeterminate, computed by the following skein relations:

\begin{align*}
(1) \quad \langle \phi \rangle &= 1 \\
(2) \quad \langle \bigcirc \bigcirc \bigcirc \rangle &= A \langle \bigcirc \bigcirc \rangle + A^{-1} \langle \bigcirc \bigcirc \bigcirc \rangle \\
(3) \quad \langle \bigcirc \bigcirc \rangle &= (-A^2 - A^{-2}) \langle \bigcirc \bigcirc \bigcirc \rangle
\end{align*}

where $\phi$ is the link with no components. Equation (2), for example, says that any time you can find three different link projections which look exactly the same except in a small disk, where they look as shown in the equation, then their brackets satisfy this equation. Of course, this means if you happen to know the brackets of the two projections on the right side, this tells you the bracket of the projection on the left side. Equation (3) is interpreted similarly and gives the effect of removing an unlinked unknot from a link.

An example of computing the bracket of a link with these rules is

\[ \langle \bigcirc \bigcirc \rangle = A \langle \bigcirc \bigcirc \rangle + A^{-1} \langle \bigcirc \bigcirc \bigcirc \rangle = A(-A^2 - A^{-2})^2 + A^{-1}(-A^2 - A^{-2}) = -A^3(-A^2 - A^{-2}). \]

This suggests the following theorem, which is proven by induction on the number of crossings.

**Theorem 1.** [Kau87] The bracket of every projection is uniquely determined by rules (1)-(3), and can be computed explicitly from them.

Generalizing the previous example,

\[ \langle \bigcirc \bigcirc \bigcirc \rangle = A \langle \bigcirc \bigcirc \bigcirc \rangle + A^{-1} \langle \bigcirc \bigcirc \bigcirc \bigcirc \rangle = -A^3 \langle \bigcirc \bigcirc \bigcirc \rangle, \]

i.e., adding a twist multiplies any bracket by $-A^3$. This is actually a positive twist, corresponding to a clockwise rotation of the framing. A negative twist would multiply the bracket by $-A^{-3}$. With this and two applications of rule (2) the reader can compute that the Kauffman bracket of the right-handed trefoil (C in Figure 1) is $A^7 + A^3 + A^{-1} - A^{-9}$.

Our interest in the Kauffman bracket is that it does not depend on the projection, but only on the unoriented framed link. This remarkable fact follows from simple calculations and the framed Reidemeister’s theorem, which says that two projections correspond to the same framed unoriented links if and only if they can be connected by a sequence of a certain finite set of local moves which are roughly Moves I, II, III of Theorem 3 in Section 2 (see [BZ85] for Reidemeister’s theorem; the framed version follows from work of Trace [Tra83]).

**Theorem 2.** [Kau87] The Kauffman bracket takes the same value on two projections of the same framed link $L$. Henceforth we refer to this value as $\langle L \rangle$. 

We would like a link invariant, rather than a framed link invariant, if for no other reason than that this is what has been studied since Kelvin (although a case can be made that framed links are a more ‘natural’ object to study: see e.g., Kirby’s theorem in Section 5). We have already seen that adding a positive or negative twist to a link changes the bracket by $-A^{\pm 3}$. It also changes the framing by one full turn. The idea is to correct the bracket so that this move leaves it unchanged, without losing the framed link invariance. To do this, consider an oriented link
projection, and label each crossing as positive or negative according to whether it looks like the first or second crossing appearing in Equation (6): i.e., according to whether one strand rotates clockwise or counterclockwise around the other. Then the \textit{writhe} of an oriented link projection is the number of positive crossings minus the number of negative crossings. For example, the writhe of E in Figure 1 is +3; that of F, its mirror image, is −3. This is an invariant of oriented framed links, which increases by one when a full twist is added. Thus if \( \mathcal{L} \) is a projection of an oriented link \( L \), the quantity

\[
(A^{-3} \text{writhe}(\mathcal{L})) \langle \mathcal{L} \rangle,
\]

where \( \langle \mathcal{L} \rangle \) is understood to be the bracket of \( \mathcal{L} \) with the orientation removed, depends only on the link \( L \).

Even better, notice that in the bracket of the trefoil, exponents increased in steps of four. This is true in general. In fact, a pretty induction on the number of crossings shows that the quantity (4) is an element of \( \mathbb{Z}[A^4,A^{-4}] \) if there are an even number of components, and is \( A^2 \) times such an element if the number of components is odd. The quantity (4), with \( t^{-1/4} \) substituted for \( A \), is called the \textit{Jones polynomial} of \( L \), or \( V_L(t) \). It sends oriented, unframed links to polynomials in \( t \) and \( t^{-1} \) (times \( t^{1/2} \) if there are an odd number of components) and satisfies the skein relations

\[
V_{\emptyset}(t) = 1
\]

\[
t^{-1}V_{\emptyset}(t) - tV_{\emptyset}(t) = (t^{1/2} - t^{-1/2})V_{\emptyset}(t)
\]

which uniquely determine it.

One can use Equations (5) and (6) to compute the Jones polynomial more efficiently than via the Kauffman bracket, but it is a bit trickier (see, e.g., Lickorish and Millet [LM88]). We should note that the Jones polynomial as it most often appears is our Jones polynomial divided by \( -t^{1/2} - t^{-1/2} \), so that its value on the unknot is 1.

It is worth doing a small computation to see what we have. From the calculations above, the Jones polynomial of the right-handed trefoil, E in Figure 1, is \( t^{1/2}(t^4 - t^2 - t - 1) \). But from the definition, the Jones polynomial of the mirror image of a link (a projection of which is gotten by reversing every crossing of a projection of the link) has the same Jones polynomial, except with \( t \) replaced by \( t^{-1} \). Thus the left-handed trefoil, F in Figure 1, has invariant \( t^{-1/2}(t^{-4} - t^{-2} - t^{-1} - 1) \). Since these are different, the two trefoils are not equivalent! Already we can see there is more afoot here than simply the fundamental group of the complement, the major tool of classical knot theory, which must be the same for both. In fact, it was not until 1914 that Dehn proved the inequivalence of these two.

The discovery of the Jones polynomial set off a rapid chain of events. It was quickly generalized by a host of people [FYH +85, PT87] to a similar, two-variable polynomial, now known as the HOMFLY polynomial (the name, formed from their initials, was unfortunately coined before the work of Przytycki and Traczyk was well known), and by Kauffman to another two-variable polynomial, the Kauffman polynomial, with very similar properties [Kau, Kau87].

Connections with statistical mechanics were quickly noticed, e.g. by Jones and Kauffman [Jon85, Kau88]. Many exactly solvable statistical mechanical systems had recently been constructed by Baxter [Bax82] from solutions of the quantum
Yang-Baxter Equation (9). Meanwhile, work in inverse scattering theory had found
solutions to the classical Yang-Baxter equation associated to Lie algebras [BD82],
and efforts to quantize these by Drinfel’d and Jimbo were just bearing fruit [Dri83,
Jim85]. This remarkable confluence resulted in the construction of quantum groups
and associated link invariants, which we sketch in the next two sections.

2. Ribbon Hopf algebras and tangle invariants

The modern view of algebraic topology is that it is about certain functors from
the geometric category of topological spaces and homotopy classes of maps, to
the algebraic category of groups and homomorphisms. The tremendous success of
this program might suggest considering other geometric categories and hoping to
find functors from them into algebraic categories. This approach may be taken as
the guiding philosophy behind the invariants discussed in this article. In fact, the
principal difference between these modern functorial invariants and the classical
ones is that here the topological spaces themselves form the morphisms.

In this section we construct the appropriate geometric category, the category of
framed tangles first described by Yetter and Turaev [Yet88, Tur89], and show that
the Kauffman bracket arises out of a functor from this category to the category of
vector spaces. We then show how to construct such a functor from the representa-
tion theory of a ribbon Hopf algebra, which we will define. In the next section, we
construct the underlying ribbon Hopf algebra for the Kauffman bracket and discuss
how to construct a large family of similar ribbon Hopf algebras. Despite the first
two paragraphs, this whole section should be accessible to someone knowing little
or no category theory.

A tangle is the image of a smooth embedding of a union of circles and intervals
into the cylinder $D \times I$, where $D$ is the unit disk in $\mathbb{C}$ and $I = [0, 1]$. The intersection
of a tangle with the boundary of the cylinder is required to be transverse, to lie
in $X \times \{0\} \cup \{1\}$, where $X$ is the $x$-axis in $D$, and to be exactly the image of
the endpoints of the intervals. Tangles are considered up to smooth isotopy of the
cylinder leaving $X \times \{0\}$ and $X \times \{1\}$ invariant. Oriented and framed tangles are
defined by analogy, but we require the framing to point in the positive $x$ direction at
the top and bottom. We define $\text{dom}(T)$ to be the number of intersection points with
$X \times \{0\}$, and $\text{codom}(T)$ to be the number of intersection points with $X \times \{1\}$ (for
oriented tangles, we must also keep track of the orientations of the points). We’ll
often speak of $T$ as a tangle ‘from’ $\text{dom}(T)$ ‘to’ $\text{codom}(T)$, making an analogy with
functions. See Figure 2 for examples.

Tangles allow two different multiplications. The composition of tangles $T_1 T_2$
is the tangle formed by putting $T_1$ on top of $T_2$, isotoping so that the endpoints
match up smoothly. This is only well-defined if $\text{dom}(T_1) = \text{codom}(T_2)$. The tensor

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{tangles.png}
\caption{Composition and tensor product of tangles}
\end{figure}
product $T_1 \otimes T_2$ is formed by putting them next to each other and treating them as a single tangle. Both multiplications are associative. See Figure 2 for examples.

In the language of category theory, let $\mathcal{T}$ be the category whose objects are nonnegative integers and whose morphisms from $n$ to $m$ are the tangles $T$ with $\text{dom}(T) = n$ and $\text{codom}(T) = m$. Composition is composition of tangles. The tensor product makes this category into what is known as a monoidal category. A monoidal category may be thought of as analogous to the category $\mathcal{V}$ whose objects are finite-dimensional vector spaces and whose morphisms are linear maps: In addition to the ordinary composition of linear maps, there is an associative multiplication, tensor product, on both the spaces and the maps.

In fact every tangle (hence every link, which is a tangle with domain and codomain 0) can be constructed out of a handful of generating tangles. Further, we can say exactly which such combinations give the same tangle. This gives a completely algebraic presentation of this monoidal category. The proof of this theorem is not difficult: It involves projections of tangles and paths between projections.

**Theorem 3.** [Tur88, Tur89, FY89] Every unoriented, framed tangle is the composition of tensor products of the five tangles $\chi, \chi, \chi, \chi, \chi$, and two such products correspond to the same tangle if and only if they can be connected by a sequence of the following moves

\[
\begin{align*}
I & \quad \begin{array}{c}
\bullet \quad \bullet
\end{array} = \end{array} & 
IV & \quad \begin{array}{c}
\begin{array}{c}
\bullet
\end{array} \quad \begin{array}{c}
\bullet
\end{array}
\end{array} = \begin{array}{c}
\begin{array}{c}
\bullet
\end{array} \quad \begin{array}{c}
\bullet
\end{array}
\end{array} \\
II & \quad \begin{array}{c}
\begin{array}{c}
\chi
\end{array} \quad \begin{array}{c}
\chi
\end{array}
\end{array} = \begin{array}{c}
\begin{array}{c}
\chi
\end{array} \quad \begin{array}{c}
\chi
\end{array}
\end{array} & 
V & \quad \begin{array}{c}
\begin{array}{c}
\chi
\end{array} \quad \begin{array}{c}
\chi
\end{array}
\end{array} = \begin{array}{c}
\begin{array}{c}
\chi
\end{array} \quad \begin{array}{c}
\chi
\end{array}
\end{array} \\
III & \quad \begin{array}{c}
\begin{array}{c}
\text{III}
\end{array} \quad \begin{array}{c}
\text{III}
\end{array}
\end{array} = \begin{array}{c}
\begin{array}{c}
\text{III}
\end{array} \quad \begin{array}{c}
\text{III}
\end{array}
\end{array} & 
VI & \quad \begin{array}{c}
\begin{array}{c}
\text{VI}
\end{array} \quad \begin{array}{c}
\text{VI}
\end{array}
\end{array} = \begin{array}{c}
\begin{array}{c}
\text{VI}
\end{array} \quad \begin{array}{c}
\text{VI}
\end{array}
\end{array} \\
& \quad \begin{array}{c}
\begin{array}{c}
\chi
\end{array} \quad \begin{array}{c}
\chi
\end{array}
\end{array} = \begin{array}{c}
\begin{array}{c}
\chi
\end{array} \quad \begin{array}{c}
\chi
\end{array}
\end{array} & 
VII & \quad \begin{array}{c}
\begin{array}{c}
\text{VII}
\end{array} \quad \begin{array}{c}
\text{VII}
\end{array}
\end{array} = \begin{array}{c}
\begin{array}{c}
\text{VII}
\end{array} \quad \begin{array}{c}
\text{VII}
\end{array}
\end{array}
\end{align*}
\]

where $T$ and $S$ are arbitrary tangles. The same is true of oriented, framed tangles, if each generator and each relation above is written with every possible consistent orientation. The same is also true in either case for unframed tangles, if one adds in relation I that both equal the identity tangle.

Much of this structure is quite natural algebraically. Moves VI and VII are just restatements of axioms of a monoidal category. More interestingly, the tangle $\chi$ is analogous to the canonical vector space map from $V \otimes W$ to $W \otimes V$, except that its square is not one. The vector space map is an example of a symmetry, a concept arising naturally in category theory (see Section 5). Instead, $\chi$ and $\chi$ satisfy the axioms for a braiding, a weakening of symmetry that comes up in studying 2-categories. In fact, the subcategory of $\mathcal{T}$ generated by $\chi, \chi$ and $\chi$ is a kind of canonical braided monoidal category: the free braided, monoidal category generated by one object. For topologically savvy readers, this category is called the braid groupoid: Restricting to tangles going from $n$ to $n$ gives the braid group mentioned in Section 1.

The tangles $\chi$ and $\chi$ are analogous to the canonical vector space map from $V \otimes V^*$ to $\mathbb{C}$ and its dual, and in particular give the category a rigidity or dual
structure, also much-studied in category theory. Leaving off Move I entirely (this corresponds to Kauffman's 'regular isotopy' of links) and orienting the tangles, we have the free rigid, braided, monoidal category on one object. There are hints that Move I is also natural in the context of $n$-categories [FY89, JS93, KV].

To understand $\mathfrak{T}$, we will consider a kind of 'representation' of it, i.e., an assignment of linear maps to the five generating tangles such that the seven relations of the previous theorem hold true, with composition interpreted as composition of linear maps and tensor product interpreted as tensor product of linear maps. This is exactly what category theorists would call a monoidal functor from $\mathfrak{T}$ to $\mathcal{V}$, and we will call it a tangle functor. In particular, the object 0 would get sent to $\mathbb{C}$ (the identity object for tensor product) and a link would get sent to a linear map from $\mathbb{C}$ to $\mathbb{C}$, which is just multiplication by a complex number. Thus we get a numerical link invariant.

Hypothesizing that the Kauffman bracket arises from a tangle functor, we see that Equations (2) and (3) are just two more equations that our five linear maps must satisfy. In fact any maps satisfying (2), (3) and IV remarkably satisfy all the relations, and straightforward calculations show they can be satisfied by operators on a two dimensional space as follows [Tur89]: Let $V$ be $\mathbb{C}^2$, and define maps

$$\langle \mathfrak{X} \rangle : V \otimes V \to V \otimes V \quad \langle \mathfrak{Y} \rangle : V \otimes V \to V \otimes V$$

$$\langle \mathfrak{A} \rangle : V \otimes V \to \mathbb{C} \quad \langle \mathfrak{I} \rangle : V \to V \quad \langle \mathfrak{U} \rangle : \mathbb{C} \to V \otimes V$$

by the following, writing the basis for $V \otimes V$ as $(1, 0) \otimes (1, 0), (1, 0) \otimes (0, 1), (0, 1) \otimes (1, 0), \text{and} (0, 1) \otimes (0, 1)$:

$$\langle \mathfrak{X} \rangle = \begin{bmatrix} A & 0 & 0 & 0 \\ 0 & 0 & A^{-1} & 0 \\ 0 & A^{-1} & A - A^{-3} & 0 \\ 0 & 0 & 0 & A \end{bmatrix} \quad \langle \mathfrak{Y} \rangle = \begin{bmatrix} A^{-1} & 0 & 0 & 0 \\ 0 & A^{-1} - A^3 & A & 0 \\ 0 & A & 0 & 0 \\ 0 & 0 & 0 & A^{-1} \end{bmatrix}$$

$$\langle \mathfrak{A} \rangle = \begin{bmatrix} 0 & A & -A^{-1} & 0 \end{bmatrix} \quad \langle \mathfrak{I} \rangle = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \langle \mathfrak{U} \rangle = \begin{bmatrix} 0 & -A & A^{-1} & 0 \end{bmatrix}^T.$$

The reader can check all the relations with nothing but elementary linear algebra (Moves III and V are best checked by applying Equation (2) first). In particular this confirms that the Kauffman bracket is a framed link invariant. Using the above, one can rewrite the Jones polynomial as a tangle functor.

We now give a set of structures and axioms on an algebra $\mathcal{A}$ which guarantee that its representation category is naturally the range of a tangle functor. That is, we will set things up so that any map sending the objects to certain representations of $\mathcal{A}$ and sending the five generators to certain canonical intertwiners (maps between representations which commute with the action of $\mathcal{A}$) automatically satisfies the seven relations. Let us proceed heuristically.

Since sequences of points are sent to a tensor product of representations, the algebra must act on such tensor products of representations. Likewise, it must act on $\mathbb{C}$, the image of the object 0. This requires a bialgebra: that is, an algebra $\mathcal{A}$, with homomorphisms $\Delta : \mathcal{A} \to \mathcal{A} \otimes \mathcal{A}$ (the coproduct) and $\epsilon : \mathcal{A} \to \mathbb{C}$ (the counit)
satisfying
\[(\Delta \otimes 1)\Delta = (1 \otimes \Delta)\Delta\]
\[(\epsilon \otimes 1)\Delta = 1 = (1 \otimes \epsilon)\Delta.\]

The reason for the terminology is that \(\epsilon\) and the adjoint of \(\Delta\) give an identity and multiplication on the dual space, making it an algebra. A bialgebra \(\mathcal{A}\) acts on \(\mathbb{C}\) by \(\epsilon : \mathcal{A} \rightarrow \mathbb{C} = \text{End}(\mathbb{C})\), and if \(\rho : \mathcal{A} \rightarrow \text{End}(V)\) and \(\tau : \mathcal{A} \rightarrow \text{End}(W)\) are representations, then \((\rho \otimes \tau)\Delta : \mathcal{A} \rightarrow \text{End}(V \otimes \text{End}(W))\) is a representation on \(V \otimes W\).

Now restricting attention to oriented tangles, recall that \(\alpha\) and \(\beta\), with various orientations, were analogous to duality in vector spaces. That is, if a boundary point with one orientation corresponds to a representation \(V\), a point with the other orientation should correspond to \(V^*\), and for example \(\alpha\) should correspond to the canonical map from \(V^* \otimes V\) to \(\mathbb{C}\). In order for \(V^*\) to be a representation, our bialgebra \(\mathcal{A}\) must be a Hopf algebra: It should have an antihomomorphism \(S : \mathcal{A} \rightarrow \mathcal{A}\) (antipode) satisfying
\[m(S \otimes 1)\Delta = m(1 \otimes S)\Delta = 1,\]
where \(m\) is the multiplication map \(m : \mathcal{A} \otimes \mathcal{A} \rightarrow \mathcal{A}\). If \(\rho : \mathcal{A} \rightarrow \text{End}(V)\) is a representation, then the dual representation \(\rho^* : \mathcal{A} \rightarrow \text{End}(V^*)\) is defined by \(\rho^*(a)v^* = v^* \circ \rho(a)\). The antipode axiom assures that there is a canonical intertwiner from \(V \otimes V^*\) to the trivial representation.

Hopf algebras have a long and distinguished history [Swe69, LR88] outside the scope of this article, but a simple example will offer motivation. Let \(G\) be a finite group and \(\mathbb{C}G\) be the group algebra. Then the maps
\[\Delta(g) = g \otimes g \quad S(g) = g^{-1} \quad \epsilon(g) = 1 \quad (\text{for } g \in G)\]
extend by linearity to make \(\mathbb{C}G\) a Hopf algebra. It is cocommutative, meaning \(\Delta\) is symmetric and thus \(\mathcal{A}^*\) is commutative.

The next piece of information is the braiding \(\lambda\), which should be an intertwiner from \(V \otimes W\) to \(W \otimes V\). It should not be the flip map, \(\rho_{VW} : v \otimes w \mapsto w \otimes v\), because then we would have \(\lambda = \lambda\) and our tangle functor would be very uninteresting! The flip map will be an intertwiner whenever \(\mathcal{A}\) is cocommutative, so we will expect interesting link invariants only from noncocommutative Hopf algebras. More precisely, if we define \(\Delta'(a) = \sigma_{\mathcal{A}\mathcal{A}}(\Delta(a))\), then we expect \(\Delta \neq \Delta'\). They shouldn’t be too different, though, since we want \(\lambda\) to behave somewhat like the flip map. It turns out the right notion is for \(\Delta\) and \(\Delta'\) to be connected by an inner automorphism. That is, following Drinfel’d [Dri85], define a quasitriangular Hopf algebra \(\mathcal{A}\) to be a Hopf algebra \(\mathcal{A}\) together with an invertible element \(R \in \mathcal{A} \otimes \mathcal{A}\), satisfying
\[\Delta'(a) = R\Delta(a)R^{-1}\]
\[(\Delta \otimes 1)R = R_{13}R_{23} \quad (1 \otimes \Delta)R = R_{13}R_{12}\]
where if \(R = \sum_i a_i \otimes b_i\), then \(R_{13} = \sum_i a_i \otimes 1 \otimes b_i\), \(R_{23} = \sum_i 1 \otimes a_i \otimes b_i\), etc. From these equations follows the Yang-Baxter equation
\[R_{12}R_{13}R_{23} = R_{23}R_{13}R_{12}\]
which will ultimately give us Move III.
Notice cocommutative Hopf algebras are always trivially quasitriangular, with 
$R = 1 \otimes 1$. Drinfel’d has shown [CP94, 4.2D][Dri87] that one can combine any Hopf 
algebra $A$ with its dual to get a larger Hopf algebra, the quantum double, which is 
quasitriangular (the ones we are interested in are almost of this form). Thus there 
are lots of them around. They also have a rich and interesting algebraic structure. 
For example, if we define $u = \sum_i S(b_i)a_i$, where $R = \sum_i a_i \otimes b_i$, then

$$S^2(a) = uu^{-1}.$$ 

This implies for example that $V$ is isomorphic as a representation to its double 
dual $V^{**}$, though not by the obvious map.

The representation theory of a quasitriangular Hopf algebra will be a rigid, 
braided, monoidal category. To get Move I requires a ribbon Hopf algebra as de-
defined by Reshetikhin and Turaev [RT90]. The ribbon structure is not as compelling 
 algebraically as the Hopf and quasitriangular structure, perhaps because Move I 
fits the categorical framework less well than the others. A ribbon Hopf algebra is a 
quasitriangular Hopf algebra with a grouplike element $G$ (i.e., $\Delta(G) = G \otimes G$ and 
$\epsilon(G) = 1$, hence $S(G) = G^{-1}$) satisfying 

$$G^{-1}uG^{-1} = S(u)$$

$$GaG^{-1} = S^2(a).$$

Given a representation $(\rho_V, V)$ and an element $x \in \text{End}(V)$, define the quantum 
trace $\text{qtr}$ by $\text{qtr}_V(x) = \text{tr}(\rho_V(G)x)$. The second condition above implies that, 
identifying $\text{End}(V)$ with $V \otimes V^*$, $\text{qtr}$ is the unique intertwiner up to scaling from this 
representation to the trivial one. That $G$ is grouplike means that the quantum trace 
is additive on direct sums and multiplicative on tensor products. Also important 
is the quantum dimension of $V$, $\text{qdim}(V) = \text{qtr}_V(1)$.

Our group algebra examples, and in fact any cocommutative Hopf algebras, are 
ribbon Hopf algebras in a trivial fashion with

$$R = 1 \otimes 1 \quad G = 1.$$ 

Of course, we do not get interesting information about links from them. For that 
we must wait until the next section.

In putting this together, we see that we have gotten more than we were originally 
looking for. Specifically, given a ribbon Hopf algebra $A$, define a labeled tangle to 
be an oriented tangle with a finite-dimensional representation of $A$ assigned to each 
component. Thus the domain and codomain are now sequences of oriented points 
labeled by representations. Assign to such a sequence a representation of $A$ by

$$\mathcal{F}((V_1, \varepsilon_1), (V_2, \varepsilon_2), \ldots, (V_n, \varepsilon_n)) = \bigotimes_{i=1}^n V_i^{\varepsilon_i},$$

where $V_i$ is a representation, $\varepsilon_i$ is + or − indicating the orientation, and $V_i^+ = V_i$, 
$V_i^- = V_i^*$.

**Theorem 4.** [RT90] For every ribbon Hopf algebra $A$ there is a monoidal functor 
from the category of labeled oriented framed tangles to that of representations and 
intertwiners of $A$. That is, there is a map $\mathcal{F}$ which assigns to each such tangle $T$ an 
intertwiner $\mathcal{F}(T) : \mathcal{F}(\text{dom}(T)) \rightarrow \mathcal{F}(\text{codom}(T))$, satisfying $\mathcal{F}(T_1T_2) = \mathcal{F}(T_1)\mathcal{F}(T_2)$ 
and $\mathcal{F}(T_1 \otimes T_2) = \mathcal{F}(T_1) \otimes \mathcal{F}(T_2)$, $\mathcal{F}(\emptyset)$ is the identity operator, and $\mathcal{F}$ of the empty 
tangle $\emptyset$ is the number 1. It is determined by its values on generating tangles, which
are shown in the following chart. Here \( G_i \) and \( R_{ij} \) are the actions of \( G \) and \( R \) on \( V_i \) and \( V_i \otimes V_j \) respectively, \( \sigma_{ij} \) is the flip map, and \( v_\alpha \) is a basis of \( V_i \), \( v_\alpha^* \) its dual basis of \( V_i^* \). Mirror images of the crossings shown get sent to their inverses.

To prove this theorem one need only check invariance under Moves I-VII. Moves II, IV, VI and VII are a straightforward calculation, Move III is the Yang-Baxter Equation (9). Moves I and V require more involved calculations.

| \( F \) of \( x^* \otimes x \) | \( x^*(x) \) | \( c \sum_\alpha v_\alpha \otimes v_\alpha^* \) |
| \( x \otimes x^* \) | \( x^*(G_i(x)) \) | \( c \sum_\alpha v_\alpha^* \otimes G_i^{-1}(v_\alpha) \) |
| \( \sigma_{ij} R_{ij}(x \otimes y) \) | \( \sigma_{ji}^* R_{ij}^*(x^* \otimes y^*) \) | \( \sigma_{ji} R_{ij}^*(x^* \otimes y) \) |
| \( \sigma_{ji}^* R_{ij}^*(x^* \otimes y) \) | \( \sigma_{ji} R_{ij}(x \otimes y) \) |

Remark 1. • There is a kind of converse of this result, a Tannaka-Krein type theorem: Every functor from \( \mathfrak{I} \) to \( \mathfrak{V} \) arises in this way from some ribbon Hopf algebra. This is proved by Majid [Maj89] for quasitriangular Hopf algebras and by the author for ribbon Hopf algebras [Saw].

• We have focused on oriented tangle functors, although currently our only interesting example is unoriented. It turns out that a self-dual representation of a ribbon Hopf algebra gives (almost) an unoriented tangle functor. As is shown in the next section, the Kauffman bracket functor essentially arises from a Hopf algebra representation corresponding to the fundamental representation of \( \mathfrak{sl}_2 \), which is self-dual.

• The use of tensor products in discussing \( \Delta \) and \( R \) is subtle in the case of most interest, when \( \mathfrak{A} \) is infinite-dimensional. All equations make sense if we require only that they hold when represented on an arbitrary finite-dimensional representation. This involves no loss of information, because in the end we are only interested in these representations [Saw].

We need to beef up our invariant in a fairly trivial fashion for future sections. Define a ribbon graph exactly as a tangle, except allow coupons, squares with a definite top, bottom and orientation, with strands allowed to intersect the coupons at distinct points on the top and bottom, as in Figure 3. Theorem 3 is still true of this larger category if we add the coupon as a generator and add relations VIII (and its mirror image) and IX in Figure 3, understood to apply for coupons with any number of lower and upper strands. Given a ribbon Hopf algebra \( \mathfrak{A} \), a labeled ribbon graph is one where the strands are labeled by representations of \( \mathfrak{A} \) as before, and any coupon with lower strands labeled \( V_1, \ldots, V_n \) and upper labeled by \( W_1, \ldots, W_m \) is labeled by an intertwiner from \( V_1 \otimes \cdots \otimes V_n \) to \( W_1 \otimes \cdots \otimes W_m \), with representations replaced by their duals if the corresponding strand is oriented down when it intersects the coupon. \( F \) extends to a functor from labeled ribbon graphs to the representation theory of \( \mathfrak{A} \), with \( F \) of a coupon labeled by \( f \) being simply \( f \). The following result summarizes important facts about \( F \).

Proposition 1. [RT90]

a. \( F \) of a tangle with a component labeled by a direct sum of representations is the direct sum of \( F \) of the tangle labeled by each summand.
An Example | Rules | Labeling
---|---|---
| | | $U \uparrow \downarrow W$

Figure 3. Coupons in ribbon graphs

b. $\mathcal{F}$ of a tangle with a component labeled by a given representation is $\mathcal{F}$ of the same tangle with the component labeled by the dual representation and given the reverse orientation.
c. $\mathcal{F}$ of a tangle with one component labeled by the trivial representation is $\mathcal{F}$ of that tangle with that component deleted.
d. $\mathcal{F}$ of a tangle with a component labeled by a tensor product of representations is $\mathcal{F}$ of that tangle with the component replaced by two parallel components, which follow the framing and are labeled by the tensor factors.
e. $\mathcal{F}$ of a tangle which can be separated by a sphere into a link and a subtangle is $\mathcal{F}$ of the link times $\mathcal{F}$ of the subtangle.
f. Let $T$ be a tangle with a component labeled by an irreducible representation $V$, and let $L$ be a link with a component labeled by $V$. If $T'$ is the tangle formed by cutting both of these components and gluing them together along the cuts (consistent with the orientations), then $\mathcal{F}(T') = \mathcal{F}(T)\mathcal{F}(L)/\mathrm{qdim}(V)$.
g. $\mathcal{F} \left( \mathcal{C}' \right) = \mathrm{qtr}(f)$.

3. Quantum groups

Up to this point, this article has been close to being self-contained: All that has been left out are computations, and with a few more precise statements and a sprinkling of hints, the industrious reader could probably reproduce them. I hope also that it has been well-motivated: that the idea of a functor from the tangle category to vector spaces is a natural thing to look for and that it might reasonably lead one to consider ribbon Hopf algebras. The same will not be true of this section.

We will see that Lie algebras offer a nice set of ribbon Hopf algebras which give trivial link information because they are cocommutative. This suggests deforming them within the set of all ribbon Hopf algebras to get interesting link invariants. If we could argue that there was a unique deformation in the space of ribbon Hopf algebras, or that some sort of quantization described it geometrically, such a construction would seem very natural. Unfortunately, these approaches give only a rough framework in which to proceed, and ultimately one is forced to write down a guess for generators and relations and prove by hand that they give a ribbon Hopf algebra. For this reason we will eschew the deep and significant mathematics in these approaches [CP94, §1-3.6.1.6.2][FT87, Dri83, Dri87] and present the algebras as if by oracle. Likewise, there is a fair amount of interesting technical machinery to prove the basic facts of these algebras [CP94, §6.4.8.1-8.3.10.1][Lus93, Lus88, Ros90a, Jim85, KR89, KR90]: We will merely sketch the easy case $sl_2$ and state the broad results in general. Ultimately, from our point of view these quantum groups are a means for extracting geometric information (link invariants) out of
quintessentially geometric objects (Lie groups), and it is certainly to be hoped that
in the near future we will have a geometric understanding of how they arise.

Recall from the previous section that the group algebra of a finite group is a
ribbon Hopf algebra, although in fact the most information it can give about a
link is the number of components. It is a little subtle deciding how to define the
analogue of the group algebra for a Lie group, but it is clear that morally the same
should be true. The best surrogate for this object turns out to be the universal
enveloping algebra of the complexified Lie algebra, $U(g)$. For the group $SU(2)$, this
is the algebra $U(sl_2)$ generated by $\{x, y, h\}$, with relations

$[h, x] = 2x$
$[h, y] = -2y$
$[x, y] = h$

where $[a, b] = ab - ba$. Its Hopf algebra structure is determined by

$\Delta(a) = a \otimes 1 + 1 \otimes a \quad \epsilon(a) = 0 \quad S(a) = -a$

where $a$ is in $\{x, y, h\}$. Its ribbon and quasitriangular structure is trivial, $R = 1 \otimes 1$
and $G = 1$. The universal enveloping algebras of the other simple Lie algebras have
similar, if more complicated, presentations [Hum72].

Given such a Lie algebra $g$, there is associated an algebra $U_s(g)$, the quantum
universal enveloping algebra of $g$, depending on a nonzero complex parameter $s$. It
will prove to be a ribbon Hopf algebra and in an appropriate sense will approach
$U(g)$ as a ribbon Hopf algebra as $s \to 1$. These quantized universal enveloping
algebras are collectively called quantum groups, though the term is sometimes used
more generally to refer to ribbon Hopf algebras, quasitriangular Hopf algebras, or
even all Hopf algebras.

$U_s(sl_2)$ is the algebra generated by $\{x, y, h\}$ subject to the relations

$[h, x] = 2x$
$[h, y] = -2y$
$[x, y] = (s^{2h} - s^{-2h})/(s^2 - s^{-2}).$

There are a number of artifices for interpreting the last equation, none entirely
satisfactory (the casual reader is encouraged to ignore the problem). Ours is as
follows: Let $U(sl_2)'$ be the completion of $U(sl_2)$ in the topology of convergence
on every finite-dimensional representation, and let the closure of the subalgebra
generated by $h$ be $\mathcal{H}$. Notice the right side of the last equation is in $\mathcal{H}$ for all
$s \neq 0$, because every finite-dimensional representation is spanned by eigenvectors
of $h$ with integer eigenvalues. Consider the algebra spanned by free products of $x,$
y and elements of $\mathcal{H}$, with the topology inherited from $\mathcal{H}$. The above equations
generate an ideal in this algebra, and the quotient by the closure of that ideal is
the algebra $U_s(sl_2)$.

This algebra is really just $U(sl_2)$ in disguise. Specifically, if we use $\bar{x}, \bar{y}, \bar{h}$
temporarily to denote the generators of $U(sl_2)$ and recall that the Casimir element
is $C = (h + 1)^2/4 + \bar{y}\bar{x}$, then the map $\varphi : U_s(sl_2) \to U(sl_2)'$ given by

$\varphi(h) = \bar{h} \quad \varphi(y) = \bar{y}$
$\varphi(x) = 4(s^{4\sqrt{C}} + s^{-4\sqrt{C}} - s^{2h-2} + s^{2-2h})/((\bar{h} - 1)^2 - 4C)(s^2 - s^{-2})^2\bar{x}$
is a homomorphism for all $s \neq 0$. Further, this homomorphism is 1-1 and has dense range if $s$ is not a root of unity (essentially because the coefficient of $\bar{x}$ is invertible) [CP94, §4.6][Jim85]. Thus up to completions $U_s(sl_2)$ and $U(sl_2)$ are isomorphic as algebras. So $U_s(sl_2)$ is semisimple, its finite-dimensional representations are the same. In particular, they have the same formal characters: That is, the traces in these representations give the same functionals on $\mathcal{H}$.

The Hopf algebra structure is given by

$$
\Delta(x) = x \otimes s^h + s^{-h} \otimes x \\
\Delta(y) = y \otimes s^h + s^{-h} \otimes y \\
S(x) = -s^2 x \\
S(y) = -s^{-2} y
$$

and the rest as in the $U(sl_2)$ case. The key observation is that these maps commute with the operator $[h, \cdot]$. From this and the fact that the formal characters are the same as for $U(sl_2)$ it follows that duals of representations and decomposition of tensor products into irreducible representations are the same (away from a root of unity) as in the classical case.

The quasitriangular structure is quite a bit trickier. Define quantum integers $[n]_q = (s^{2n} - s^{-2n})/(s^2 - s^{-2})$. Then

$$
R = s^{(h \otimes h)} \sum_{n=0}^{\infty} \frac{s^n(n+1)(1-s^{-4})^n}{[n]_q[n-1]_q \cdots [1]_q} x^n \otimes y^n \\
G = s^{2h}.
$$

The infinite sum converges on every pair of representations because only finitely many terms are nonzero, as follows from the next paragraph. Thus it is in the completion of $U_s(sl_2) \otimes U_s(sl_2)$ with respect to this topology.

If you believe that all the axioms of a ribbon Hopf algebra are satisfied by the algebra above, you should be wondering what the tangle functor looks like. First we must understand the representation theory explicitly. For this recall that $sl_2$ has an $n$-dimensional representation for each $n > 0$, with basis $\{v_i\}_{i=1}^n$ such that

$$
h v_i = (n - 2i + 1) v_i \\
x v_i = (i - 1) v_{i-1} \\
y v_i = (n - i) v_{i+1}
$$

and that the action of $h$ is independent of $s$ (by the isomorphism). It is easy to reconstruct the representation $V_n$ of $U_s(sl_2)$ with basis $\{v_i\}_{i=1}^n$ such that

$$
h v_i = (n - 2i + 1) v_i \\
x v_i = [i - 1]_q v_{i-1} \\
y v_i = [n - i]_q v_{i+1}.
$$

One can now compute the action of $R$ on $V_2 \otimes V_2$, and composing with the flip map gives

$$
\begin{pmatrix}
\bar{\alpha}
\end{pmatrix} \rightarrow 
\begin{bmatrix}
s & 0 & 0 & 0 \\
0 & 0 & s^{-1} & 0 \\
0 & s^{-1} & s - s^{-3} & 0 \\
0 & 0 & 0 & s
\end{bmatrix}.
$$

Identifying $V$ with $V^*$ by the isomorphism $\alpha : v_1 \mapsto sv_2^*$, $\alpha : v_2 \mapsto -s^{-1} v_1^*$ gives

$$
\begin{pmatrix}
m
\end{pmatrix} = \begin{bmatrix}0 & -s & s^{-1} & 0\end{bmatrix}, \\
\begin{pmatrix}
p
\end{pmatrix} = \begin{bmatrix}0 & s & -s^{-1} & 0\end{bmatrix}.
with the other generating tangles determined by these. Notice this is almost an
unoriented tangle functor—in fact the Kauffman bracket functor, Equations (7)
and (8), with $A = s$—except for an annoying minus sign in $\tilde{\alpha}$. The link invariant
will actually be the Kauffman bracket times $(-1)^w$, where $w$ is the total winding
number of the link projection ($(-1)^w$ is a framed, unoriented link invariant!). In
general any irreducible self-dual representation gives, on choosing such an $\alpha$, either
an unoriented functor or a functor which differs from one by a minus sign in $\tilde{\alpha}$.
Kirillov and Reshetikhin [KR89] nicely fix this problem for $sl_2$ by constructing
$U_q(sl_2)$ as a deformation of $U(sl_2)$ with a nonstandard quasitriangular structure.

We have only described the functor with everything labeled by the two-dimen-
sional representation, but we have really constructed much more than that. There
is enough information above to write down the full theory for all representations,
but no neat skein-theory description has been found for the corresponding invariant.

It is now clearer where the skein relation in Equation (2) is coming from. $V_2 \otimes
V_2$ breaks up as the sum of two irreducible representations, the trivial and $V_3$,
and thus the space of intertwiners on it is two-dimensional. The two fragments
pictured on the right side of Equation (2), interpreted as tangles, correspond under
$F$ to the identity intertwiner and a multiple of the projection onto the trivial
subrepresentation. They span the space of intertwiners, which includes $\mathfrak{A}$.

As for other Lie algebras, the story is essentially the same. A similar, but more
complicated presentation of $U_q(g)$ can be given and shown in the same sense to
be isomorphic to $U(g)$ (though in general the isomorphism does not admit explicit
formulae as in the $sl_2$ case). The ribbon Hopf structure can all be written down
explicitly, and the representation theory is ‘the same’ as for the original Lie algebra
except at roots of unity [CP94, §8.3,10.1]. The link invariants coming from the funda-
mental representation of $sl_n$ are particular values of the two-variable HOMFLY
polynomial. The polynomials for $B_n$, $C_n$, and $D_n$ at the fundamental representa-
tion are all special values of the two-variable Kauffman invariant [Tur89, Res88].
Both of these polynomials can be computed by skein-theoretic algorithms similar
to that for the Jones polynomial. There is also a skein-theory algorithm for the $G_2$
invariant at the fundamental representation involving trivalent graphs [Kup94].

Remark 2. A word on notation. It is customary to speak of the quantum groups
as $U_q(g)$, where $q$ is, depending on the author, $s^2$ [CP94, Kas94, Lus93, Jim85] or
$s^4$ [Res88, Ros90b, Dri87, KM91, RT90, RT91]. Of course, this requires choosing a
second or fourth root of $q$ to write the $R$ matrix. It is also common to call this vari-
able $t$, or to call the variable in the Jones polynomial $q$, although these are inverses
of each other with the usual conventions. Other minor variations exist on the defi-
nition of $U_q(g)$. We have tried to remain internally consistent and consistent with
well-established conventions, such as the skein relations for the Jones polynomial.

4. MODULAR HOPF ALGEBRAS

What happens to the quantum group at a root of unity? The answer is, the
algebra ceases to be isomorphic to the unquantized algebra and in fact ceases to
be semisimple. The ribbon Hopf structure and some of the representation theory
survive this collapse however and emerge with subtle properties that allow the link
invariant to be extended to a 3-manifold invariant and what’s called a topological
quantum field theory. This is the subject of the next four sections. We begin with
a precise look at the situation for $sl_2$. 
When \( s \) is a root of unity, the representations of \( U_s(\mathfrak{sl}_2) \) given in the previous section are still representations, but may not all be irreducible. It can be shown that irreducibility is equivalent to \( v_1 \) being the unique \( h \)-eigenvector on which \( x \) acts as 0. Thus if we let \( r \) be the least natural number such that \( s^{4r} = 1 \) (so that \( s \) is a primitive \( 4r \)-th or \( 2r \)-th root of unity), we have \([r]_q = 0\), \([n]_q \neq 0\) for \( n < r \), and thus \( V_n \) is still irreducible for \( n \leq r \), but is not for larger \( n \), by Equation (13). Recalling that the quantum dimension of a representation is the trace of \( G \) in that representation, notice that the quantum dimension of \( V_n \) is \([n]_q\) and thus is nonzero for \( n < r \) and equal to zero for \( n = r \). It turns out that irreducible representations with quantum dimension zero have an important property: If one takes any representation in the ideal generated by them, i.e., a direct sum of tensor products of these with other representations, then any intertwiner from this representation to itself will have quantum trace zero. But this implies that any link labeled by such a representation has invariant zero. It is easy to check that the tensor product \( V_n \otimes V_m \) for \( n, m < r \) is a sum of representations \( V_k \) for \( k < l \) plus a representation in this ideal. Since for purposes of knot theory the representations in this ideal are irrelevant, it makes sense to throw this trivial-trace representation out and define the truncated tensor product to be the rest of the tensor product. With this new tensor product, direct sums of representations \( V_n \) for \( n < r \) still form a rigid, braided, tensor category and give a tangle functor. But now there are only finitely many essentially different labels.

The same situation applies to a general quantum group, though to state it precisely requires more knowledge of Lie algebras than we have needed yet (see e.g. [Hum72]). If \( s \) is a primitive \( 4r \)-th root of unity, or \( 2r \)-th or \( r \)-th for \( r \) odd, then the representation of highest weight \( \lambda \) is still an irreducible representation with the correct character when \( (\lambda + \rho, \theta) \leq r \), where \( \theta \) is the highest root (the unique long root in the Weyl chamber), \( \rho \) is half the sum of the positive roots, and \((\cdot, \cdot)\) is the invariant inner product on the Lie algebra, normalized so that long roots have square length 2 (this differs from the usual convention for quantum groups, but makes the results simpler to state and is consistent with the convention for affine Lie algebras [KP84]). This representation has nonzero quantum dimension exactly when the inequality is strict. Further, the tensor product of two representations in this set is a direct sum of representations in the set and a representation in the ideal generated by those representations where the inequality is an equality, which all have quantum dimension 0. Again, this gives a ‘truncated’ tangle functor [AP95]. Of course, this is only interesting if more than the trivial representation is included, so we usually restrict to \( r \) such that \( r > (\rho, \theta) \). The quantity \( r - (\rho, \theta) \) is the \( k \) occurring in the Chern-Simons action in the introduction.

The other useful property of quantum groups at roots of unity has to do with the value of the invariant on the Hopf link, \( G \) in Figure 1. Let \( H_{i,j} \) be the value of \( \mathcal{F} \) on the Hopf link labeled by representations \( \lambda_i \) and \( \lambda_j \). For \( U_s(\mathfrak{sl}_2) \), a straightforward calculation shows that \( H_{i,j} = [ij]_q \), where \( [i] \) is the \( i \)-dimensional representation. At a root of unity, if we restrict to our preferred representations, these numbers form a matrix, and another calculation shows that this matrix is nonsingular when \( s \) is a primitive \( 4r \)-th root of unity [RT91]. It is singular for the other roots of unity, though primitive \( 2r \)-th roots with \( r \) odd can be patched up to give a kind of TQFT [BHM92]. Kirillov [Kir] has shown, by relating the question to work of Kac and Petersen [KP84], that the same matrix for an arbitrary quantum group...
is nonsingular when \( s \) is again a 4\( r \)th root of unity. What happens at other roots of unity is not yet known and does not apparently correspond to a physical field theory.

This information is encapsulated in the definition of a modular Hopf algebra (see [RT91], in this simplified form [Tur92, TW93]).

**Definition 1.** A modular Hopf algebra is a ribbon Hopf algebra \( A \), together with a finite collection of irreducible representations \( \lambda_1, \ldots, \lambda_k \), including the trivial representation, closed under duals, and having \( \text{qdim}(\lambda_i) \neq 0 \), satisfying

- \( \lambda_i \otimes \lambda_j = \bigoplus_{m=1}^{k} (N_{i,j}^m \lambda_m) \oplus \eta_{i,j} \), where \( N_{i,j}^m \) is a multiplicity and \( \eta_{i,j} \) is a representation on which \( \text{qtr} \) is zero on all intertwiners.
- The matrix \( H_{i,j} \), for \( 1 \leq i, j \leq k \) is nonsingular.

The above results can be formulated as “\( U_q(\mathfrak{g}) \) is a modular Hopf algebra when \( q \) is a root of unity,” but only with the convention \( q = s^4 \) of, e.g. [RT90, RT91]. It is incorrect even for \( \mathfrak{sl}_2 \) with the more prevalent \( q = s^2 \) convention, though the second formulation has occurred frequently in the literature [TW93, And92].

**Remark 3.** That the quantum dimension of each \( \lambda_i \) is nonzero actually follows from the second condition of the definition.

We will now derive the key facts about a modular Hopf algebra that we will need later on. Let \( R \) be the commutative algebra over \( \mathbb{C} \) spanned by \( \lambda_1, \ldots, \lambda_k \) with \( \lambda_i \lambda_j = \sum_m N_{i,j}^m \lambda_m \). We call \( R \) the truncated representation ring of the modular Hopf algebra.

We can extend \( \text{qdim} \) by linearity to a homomorphism from \( R \) to \( \mathbb{C} \). More generally, if \( T \) is a ribbon tangle which is labeled except for one closed component \( x \) without coupons and \( a = \sum_i c_i \lambda_i \in R \), we can define \( T \) with \( x \) labeled by \( a \) to be the formal sum \( T_a = \sum_i c_i T_{\lambda_i} \), where \( T_{\lambda_i} \) is \( T \) with \( x \) labeled by \( \lambda_i \). Then define the value of the invariant

\[
F(T_a) \overset{\text{def}}{=} \sum_{i=1}^{k} c_i F(T_{\lambda_i}).
\]

This is consistent with the additivity of the invariant on direct sums.

It is natural to ask if \( R \) is semisimple: that is, is it spanned by minimal idempotents? In fact the dual basis to \( \{ \lambda_i \} \) with respect to the nondegenerate, symmetric pairing given by \( \{ H_{i,j} \} \) consists of such minimal idempotents. Minimal idempotents give homomorphisms to \( \mathbb{C} \), and the idempotent dual to the trivial representation gives the homomorphism \( \text{qdim}(\cdot) \). This minimal idempotent, which is the source of the 3-manifold invariant, is computed explicitly in the following proposition.

**Proposition 2.** [Saw] Let

\[
\omega = \sum_{i=1}^{k} \text{qdim}(\lambda_i) \lambda_i.
\]

a. \( a \omega = \text{qdim}(a) \omega \) for all \( a \in R \).

b. The Hopf link labeled by \( \lambda_i \) and \( \omega \) has nonzero \( F \) if and only if \( \lambda_i \) is trivial (in particular \( \text{qdim}(\omega) \neq 0 \)).
5. AXIOMATIC TOPOLOGICAL QUANTUM FIELD THEORY

The fundamental idea of the first three sections was to get a detailed combinatorial description of an interesting geometric category \((\Sigma)\) and use it to find functors to the category of vector spaces and linear maps. We do the same thing with the cobordism category and call it a topological quantum field theory, or TQFT. Here we obtain numerical invariants not of (framed) links, but of (biframed) 3-manifolds. The work is remarkably parallel, the key difference being that this category is not presented as straightforwardly: It is given indirectly, in terms of links. Thus instead of TQFTs being associated to a specific kind of algebra, we construct them from a specific kind of link invariant: one arising from a modular Hopf algebra.

Let \(\Sigma_1\) and \(\Sigma_2\) be smooth oriented \(d-1\)-manifolds. A dimension \(d\) cobordism \(m\) with domain \(\Sigma_1\) and codomain \(\Sigma_2\) is up to diffeomorphism a triple \((M, f_1, f_2)\), where \(M\) is an oriented smooth \(d\)-dimensional manifold with boundary and \(f_1\) and \(f_2\) are orientation-preserving 1-1 diffeomorphisms from \(\Sigma_1^*\) and \(\Sigma_2\) respectively to \(\partial M\). Here \(\Sigma_1^*\) is \(\Sigma_1\) with the opposite orientation, and we require of \(f_1\) and \(f_2\) that \(\partial M\) be the disjoint union of their ranges: That is, \(\partial M\) is parametrized by \(\Sigma_1^* \cup \Sigma_2\).

By “up to diffeomorphism” we mean that two triples \((M, f_1, f_2)\) and \((M', f'_1, f'_2)\) are the same morphism if there is a diffeomorphism \(F: M_1 \to M_2\) with \(f'_1 = Ff_1\) and \(f'_2 = Ff_2\). Call the image of \(f_1\) the ‘incoming boundary’ and that of \(f_2\) the ‘outgoing boundary’. See Figure 4 for an illustration. Cobordisms are reminiscent of the tangles of Section 2, and recalling the category \(\mathcal{V}\) of vector spaces and linear maps discussed in that section, we will again think of the \(\Sigma_i\) as being analogous to vector spaces and cobordisms as being analogous to linear maps.

In this vein, the analogue of composition of linear maps is gluing of cobordisms. That is, if \(m = (M, f_1, f_2)\) and \(m' = (M', f'_1, f'_2)\), define \(m \circ m' = (M \cup f'_1 f_2^{-1}, f_1, f_2)\), where \(M \cup f'_1 f_2^{-1}\) is the manifold formed by identifying points in \(\partial M'\) with \(\partial M\) via the orientation reversing map \(f'_1 f_2^{-1}\). This composition is associative, and its identity is clearly

\[1_\Sigma = (\Sigma \times I, \text{id}_0, \text{id}_1),\]

where \(I\) is the unit interval and \(\text{id}_0 : \Sigma^* \to \Sigma \times \{0\}\) and \(\text{id}_1 : \Sigma \to \Sigma \times \{1\}\) are the identity maps.

Tensor product of vector spaces is analogous to disjoint union \(\Sigma_1 \cup \Sigma_2\); \(\mathbb{C}\), the identity for tensor product, corresponds to 0, the empty \(d-1\)-manifold. Corresponding to tensor product of linear maps, we have the associative operation \(m \otimes m' = (M \cup M', f_1 \cup f'_1, f_2 \cup f'_2)\). The empty \(d\)-manifold \(\emptyset\) acts like the number 1 as the identity for this product.

\[\text{Figure 4. A typical cobordism with parametrizations } f_1 \text{ and } f_2\]
Of course $V \otimes W$ and $W \otimes V$ are the same vector space, in the sense that there is a canonical cobordism map between them $\sigma_{VW} : v \otimes w \mapsto w \otimes v$. Likewise there is a canonical cobordism from $\Sigma_1 \cup \Sigma_2$ to $\Sigma_2 \cup \Sigma_1$, namely, the cobordism $\sigma_{\Sigma_1, \Sigma_2} = ((\Sigma_1 \cup \Sigma_2) \times I, f_0, f_1)$, with

$$f_0 : \Sigma_1 \cup \Sigma_2 \to (\Sigma_1 \cup \Sigma_2) \times \{0\} \quad f_1 : \Sigma_2 \cup \Sigma_1 \to (\Sigma_1 \cup \Sigma_2) \times \{1\}$$

being the identity and the order-reversing map respectively. Notice here cobordisms behave a little better than tangles did, because the analogue of the symmetry for tangles was the braiding, whose square was not one.

No mathematician, and certainly no category theorist, would let such a good analogy go unnamed. The definition of cobordisms is designed to make them the morphisms of a category, which we will call $\mathcal{C}$, and all the additional structures we discussed are just to say that $\mathcal{C}$ and $\mathcal{B}$ (the category of vector spaces) are symmetric, monoidal or tensor categories (there are some axioms to check, which are trivial in both cases. See Mac Lane [Mac71]).

The point of this is

**Definition 2.** [Ati89, Ati90a] A $d$-dimensional axiomatic topological quantum field theory, or TQFT, is a functor $Z$ of tensor categories from $\mathcal{C}$ to $\mathcal{B}$. That is, a map $Z$ which sends each oriented $d-1$-manifold to a finite-dimensional vector space $Z(\Sigma)$ and each cobordism $m$ from $\Sigma_1$ to $\Sigma_2$ to a linear map $Z(m) : Z(\Sigma_1) \to Z(\Sigma_2)$, such that $Z(1_{\Sigma}) = 1_{Z(\Sigma)}$, $Z(\Sigma_1 \cup \Sigma_2) = Z(\Sigma_1) \otimes Z(\Sigma_2)$, $Z(0) = \mathbb{C}$, $Z(m_1 \cup m_2) = Z(m_1) \otimes Z(m_2)$, $Z(\emptyset) = 1$, and $Z(c)$ is the flip map.

Notice that a closed $d$-manifold is a cobordism from the empty $d-1$-manifold to itself and thus gets sent to an operator on $\mathbb{C}$, or in other words a number. So one thing a TQFT gives is a numerical $d$-manifold invariant. There is a straightforward algorithm for computing it from a Morse function: If you can compute the operators assigned to cobordisms for each type of critical point, you can multiply them together to get the invariant of the closed manifold. Also, it follows that the mapping class group of any $d-1$-manifold acts on the vector space associated to that manifold. Thus these contain a lot of topological information [Tur94, §III].

The definition of a TQFT is modeled on what we expect a topological quantum field theory to yield. For the passage from the physics to these axioms, see [Ati90a] or Axelrod [Axe91].

Let us now fix $d = 3$. We will not actually construct a functor from $\mathcal{C}$, but from a slightly larger category $\mathfrak{C}$ of biframed (also called 2-framed) cobordisms, which we define in the next paragraph. The reason for this lies buried in the subtleties of the Chern-Simons path integral, from which these examples spring, where a choice of biframing is necessary to regularize certain integrals. The casual reader may safely ignore the technicalities of biframings and view them as playing a role very similar to that of framings on links for link invariants.

A biframing on a closed 3-manifold is a trivialization of $TM \oplus TM$, up to isotopy. Atiyah shows [Ati90b] that a given trivialization extends to a 4-manifold bounded by that 3-manifold exactly when the 4-manifold has a certain corresponding signature, which is a complete invariant of the biframing. A biframing on a manifold with boundary $\Sigma$ restricts to a 1-1 map from $T\Sigma \oplus T\Sigma$ to $\mathbb{R}^6$. Call a biframed 2-manifold one equipped with such a map. A biframed cobordism is a 3-cobordism together with a trivialization up to isotopy fixing the boundary, and its domain and codomain are thus biframed 2-manifolds. Biframed cobordisms form a tensor category $\mathfrak{FC}$.
exactly as before, with the observation that if \( \Sigma \) is a biframed 2-manifold, \( \Sigma^* \) inherits its biframing and \( 1_\Sigma \) and \( \phi_{\Sigma, F} \) have canonical biframings.

We can make this category simpler to work with in a number of ways. First, two biframed 2-manifolds of the same genus can be connected by an invertible biframed cobordism. Thus their vector spaces are isomorphic, and we only need to pick a vector space for each one (this reduction is what category theorists call ‘skeletonization’. The category of tangles is really a skeletonization of a larger category with a geometrically more natural definition). The fact that \( \mathfrak{FC} \) has a duality structure similar to that of tangles (view \( \Sigma \) as a morphism from \( \Sigma^* \cup \Sigma \) to 0) means we really do not have to distinguish between domain and codomain. Finally, the value of the functor on disconnected cobordisms is clearly determined by its value on connected ones.

To state the simplified version, for each \( g \) choose a representative biframed genus \( g \) surface \( \Sigma_g \) and a biframed cobordism \( \partial_g : \emptyset \rightarrow \Sigma_g \cup \Sigma_g \) whose underlying manifold is \( \Sigma_g \times I \) and which is symmetric in the sense that \( \phi_{\Sigma_g \cup \Sigma_g} \partial_g = \partial_g \). Also let \( \partial_g \) be a biframed cobordism such that \( (\partial_g \otimes 1)(1 \otimes \partial_g) = 1 = (1 \otimes \partial_g)(\partial_g \otimes 1) \).

**Theorem 5.** [Saw] For each \( g \), let \( Z(\Sigma_g) \) be a finite-dimensional vector space; for each biframed cobordism \( m : \bigsqcup_{i=1}^n \Sigma_{g_i} \rightarrow \emptyset \), let \( Z(m) : \bigotimes_{i=1}^n Z(\Sigma_{g_i}) \rightarrow \mathbb{C} \) be a linear functional (in particular, \( Z \) sends closed manifolds to functionals on \( \mathbb{C} \), i.e., elements of \( \mathbb{C} \)). Then \( Z \) extends uniquely to a biframed TQFT, if and only if the following hold:

a. **Nondegeneracy:** \( Z(\partial_g) \) is a symmetric nondegenerate pairing on \( Z(\Sigma_g) \).

b. **Symmetry:** If \( m \) has domain \( \bigsqcup_{i=1}^n \Sigma_{g_i} \), and \( m' \) with domain \( \bigsqcup_{i=1}^n \Sigma_{g_{i', i}} \) for some permutation \( \sigma \) is the same manifold as \( m \) with the same parameterization written in a different order, then \( Z(m) = Z(m')P_{\sigma} \), where \( P_{\sigma} \) is the map on the appropriate tensor product of vector spaces which permutes the tensor factors.

c. **Sewing:** Suppose \( m \) has domain \( \bigsqcup_{i=1}^m \Sigma_{g_i} \), and \( n \) has domain \( \bigsqcup_{i=1}^n \Sigma_{g_{i'}} \) with \( g_m = g_1 \). We can form the biframed manifold \( m \cup_s n \), with domain \( \bigsqcup_{i=1}^{m+n} \Sigma_{g_{i'}} \cup \bigsqcup_{i=2}^{m+n} \Sigma_{g_{i'}} \), by composing with \( \partial_g \) along these boundary components (Figure 5).

Then
\[
Z(m \cup_s n) = Z(m) \otimes Z(n) \circ \alpha
\]
where \( \alpha \) is the canonical map sending \( v_1 \otimes \cdots \otimes v_{m-1} \otimes w_2 \otimes \cdots \otimes w_n \) to \( \sum_i v_1 \otimes \cdots \otimes v_{m-1} \otimes a_i \otimes b_i \otimes w_2 \otimes \cdots \otimes w_n \), with \( a_i \) and \( b_i \) dual bases of \( Z(\Sigma_{g_{i+1}}) \) with respect to the pairing \( Z(\partial_g) \).

d. **Mending:** Suppose \( m \) has domain \( \bigsqcup_{i=1}^m \Sigma_{g_i} \), with \( g_1 = g_2 \). We can form \( m_m \) by composing with \( \partial_g \) along the first two boundary components (Figure 5).

Then
\[
Z(m_m) = Z(m) \circ \alpha
\]
where \( \alpha \) is the canonical map as above.

![Figure 5. Examples of sewing and mending](image-url)
Remark 4. • The distinction between sewing and mending may seem arbitrary. However, from the point of view of the construction in Section 7 they are quite different. In fact, a careful examination of the proof of Theorem 7 shows that if we dropped the mending axiom and the requirement that the vector spaces be finite-dimensional, all 3-manifold invariants would arise from TQFTs!
• There is one obvious piece of structure on $C$ that we are leaving off: Reversing the orientation on a cobordism from $\Sigma$ to $\Sigma'$ gives a cobordism from $\Sigma'$ to $\Sigma$. This is analogous to Hilbert space adjoint. It thus seems natural to replace $\mathcal{Y}$ with the category of finite-dimensional Hilbert spaces and ask that this structure too be preserved. This is called a unitary TQFT and is the more natural one from the point of view of physics. It turns out that, while we will construct a TQFT for any quantum group and any primitive 4th root of unity, it is only unitary for $s = e^{\pi i/2r}$. This is exactly the value of $s$ corresponding to Witten’s geometric construction.
• Of course, an ordinary framing on a closed 3-manifold determines a biframing, so in the end we will have an invariant of framed 3-manifolds as well.

6. Surgery and cobordism

We would like to have an algebraic description of the category $\mathfrak{FC}$ of framed cobordisms in terms of generators and relations. The idea is to define a TQFT by the image of the generating morphisms and prove that it is a TQFT by confirming that it preserves the relations, exactly as we did with tangles. A natural choice is Morse theory. Here the generators are the morphisms attaching handles and the relations are those given by Cerf theory [Cer70]. This is the approach taken in [Wal], but it involves a lot of detail checking, because there are a lot of ways to attach a handle.

Surgery on links meets our present need much better. Since surgery describes 3-manifolds in terms of links and we are constructing TQFTs out of link invariants, we will find many of the details fall into place. Of course, surgery does not apparently handle biframing or manifolds with boundary, so this section is devoted to extending it to these situations. We begin with a review of (integer) surgery.

Consider a framed unoriented link $L$ in $S^3$. Let $T$ be a tubular neighborhood of $L$. Each component of $\partial T$ has a natural meridian, which bounds a disk in $T$. It also has a natural longitude, the pushoff of $L$ in the direction of the framing. Remove each component of $L$ and glue it back in by a map sending the meridian to the longitude and the longitude to minus the meridian (you must choose orientations for these curves, but the result does not depend on this choice). The manifold one obtains is called $M_L$, the result of surgery on $L$. For example, surgery on the 0-framed unknot gives $S^1 \times S^2$, and surgery on a $\pm 1$ framed unknot gives $S^3$ again.

It turns out, by a theorem of Lickorish [Lic62], that every closed, compact, connected, oriented 3-manifold admits such a presentation. Further, by a theorem of Kirby [Kir78], two links present the same 3-manifold if and only if they can be related by a sequence of the moves in Figure 6 and their mirror images, with any number of strands understood to pass through the pictured unknot. Actually, Kirby’s original version involved two “nonlocal” moves: Fenn and Rourke [FR79] reduced them to this move.
Kirby’s proof takes place in four dimensions. He views $S^3$ as bounding $B^4$ and interprets $L$ as defining a 4-manifold by attaching 2-handles to tubular neighborhoods of the components. The resulting 4-manifold will have boundary $M_L$. Thus surgery might better be viewed as resulting in a 3-manifold together with a choice of 4-manifold for it to bound.

This is perfect for us, because this is just what we saw in the previous section determines a biframing. Of course we need a refinement of the Kirby move which preserves the signature of the 4-manifold but is still powerful enough to connect all links giving the same signature. This task at first sounds daunting but is in fact fairly simple. The key observation is that, since each component of $L$ corresponds to adding a 2-handle to the 4-manifold, the second relative cohomology of the 4-manifold $M$ has a generator for each handle, supported in that handle. Further, the intersection form of two generators is the linking number of the corresponding components (the self-intersection is the self-linking number, determined by the framing). Thus the signature of the 4-manifold is exactly the signature of the matrix of linking numbers of $L$!

A straightforward calculation shows that the Kirby move changes the signature of the linking matrix by 1. Thus the two biframed Kirby moves pictured in Figure 7, each a composition of two Kirby moves, do not change the signature. On the other hand, if two links represent the same 3-manifold and have the same signature, one can convert the sequence of ordinary Kirby moves connecting them to a sequence of biframed Kirby moves.

**Theorem 6.** [Saw] Closed, compact, oriented, biframed 3-manifolds are in one-to-one correspondence with equivalence classes of biframed unoriented links in $S^3$ modulo moves I and II in Figure 7 and their mirror images.

The cobordism category is not much harder. For each surface $\Sigma_g$ choose a biframed handlebody which it bounds, $H_g$. To represent pictorially an embedding of $H_g$ into $S^3$, choose a set of generators for the fundamental group of the handlebody, $\{c_i\}$ for $1 \leq i \leq g$. Represent an embedding of $H_g$ into $S^3$ as in Figure 8. The embedding is recovered by thickening the framed graph to a handlebody and identifying $H_g$ with the thickening by a map which sends $c_i$ to the boundary above
Presenting handlebody embeddings
Boundaries identified by $\partial_g$

**Figure 8.** A pictorial description of the cobordism category

the $i$th circle on the graph. Choose $\partial_g$ so that the identification of $\Sigma_g$ with $\Sigma_g^*$ sends each $c_i$ to a meridian intersecting it and vice versa, as illustrated for genus two in Figure 8.

Now if $m : \bigcup_{i=1}^n \Sigma_{g_i} \to \emptyset$ is a connected cobordism, consider the closed manifold $M$ formed by gluing $\bigcup_{i=1}^n H_{g_i}$ to $m$ along the boundary by the parametrization, and consider the resulting embedding $N$ of $\bigcup_{i=1}^n H_{g_i}$ into $M$. This pair determines $m$ uniquely. $M$ can be presented as $M_L$ for some link $L$, and the embedding $N$ can be isotoped so as not to intersect the embedded tori coming from surgery on $L$. So $N$ corresponds to an embedding of $\bigcup_{i=1}^n H_{g_i}$ into $S^3$ which does not intersect $L$. Thus $m$ is determined by a pair $(N, L)$, where $L$ is an unoriented biframed link and $N$ is an embedding of handlebodies into $S^3$ which does not intersect $L$. Specifically, $m$ is obtained by doing surgery on $L$, removing the image of the interiors of the handlebodies, and parametrizing the resulting boundary by $N$ restricted to the boundary. We call $(N, L)$ a presentation of $m$ and in general refer to the cobordism presented by $(N, L)$ as $[N, L]$.

In [Rob] a set of Kirby-type moves is given for manifolds with boundary. From this it follows that two presentations give the same cobordism if and only if they can be connected by isotopy and a sequence of Kirby moves (Figure 6) and their mirror image. The framed unknot in Figure 6 represents a component of $L$, and the strands passing through it represent pieces of $L$ or $N$. We thus have a purely combinatorial description of cobordisms.

In fact, $(N, L)$ determines a biframed cobordism, and Figure 7 connects all equivalent presentations of a biframed cobordism. Specifically, having chosen a biframing on $H_{g_i}$, biframings on $[N, L]$ are in one-to-one correspondence with biframings on $M_L$, which are preserved by the biframed Kirby moves.

Our last goal is to give a presentation of a cobordism obtained by sewing or mending, in terms of the presentation of the manifold(s) to be sewn or mended. The following proposition is not hard to show from the biframed Kirby moves.

**Proposition 3.** [Saw] Any biframed cobordism can be written as $[N, L]$, where $N$ is in standard position in the sense that it can be projected with no self-crossings, as illustrated in Figure 9.

Thus we may assume the presentations to be sewn or mended are in standard form. It is more clear and convenient to describe the algorithm pictorially, as in Figures 9. The corresponding picture applies for any genus and any number of strands passing through the handles, but assumes the $\partial_g$ given in Figure 8.

The algorithm for sewing follows straightforwardly from the definition, but the algorithm for mending needs a little justification. We first note that, to form the
mending \([N, L]_g\), we might just as well form the mending \([N, \emptyset]_m\) and then do surgery on the image of \(L\) in this new manifold. But now because \(N\) is in standard position, \([N, \emptyset]_m\) is simply \(S^1 \times S^2\) (perhaps with some handlebodies removed), and the picture we get is exactly the right side of Figure 9.

7. Constructing the TQFT

We are now ready to construct a TQFT. For this we need to associate vector spaces to surfaces. We first associate excessively large vector spaces to each \(\Sigma_g\), spanned by labeled ribbon graphs in \(H_g\). Then we give a pairing on this corresponding to \(\varepsilon_g\) and quotient by the null space to force the pairing to be nondegenerate. The value of the invariant on cobordisms is then essentially forced upon us.

Given a biframed handlebody \(H_g\), a closed labeled ribbon graph in \(H_g\) is defined as in Section 2, except that the graph is embedded in \(H_g\) and equivalence is by a smooth isotopy fixing the boundary.

Let \(V_g\) be the vector space of all formal linear combinations of closed labeled ribbon graphs in \(H_g\). For each presentation \((N, L)\), where \(N\) is an embedding of \(\bigcup_{j=1}^n H_{g_j}\) into \(S^3\), we define a map \(f(N, L): \bigotimes_{j=1}^n V_{g_j} \to \mathbb{C}\), as follows. If \(h_j\) is a closed labeled ribbon graph in \(H_{g_j}\) for \(1 \leq j \leq n\), embed \(\bigcup_{j=1}^n h_j\) into \(S^3\) via \(N\) and label each component of \(L\) by \(\Omega = q\text{dim}(\omega)\), with \(\omega\) as in Proposition 2. Let \(K = q\text{dim}(\Omega)\), and define \(f(N, L)(\bigotimes h_j)\) to be \(K^{-1}\) times \(F\) of the resulting graph.

**Proposition 4.**

a. If Moves I and II of Figure 7 represent closed labeled ribbon graphs in \(S^3\) with the framed unknots pictured labeled by \(\Omega\), the values of \(F\) on both sides of the equal signs agree.

b. The function \(f(N, L)\) depends only on \([N, L]\), and not on the presentation. Thus we will speak henceforth of \(f[N, L]\).

**Proof.**

a. For Move II, use Proposition 1 and Figure 10. The first step is by parts (a) and (d), the second by part (f), the third by Proposition (2a), and the last by (a) and (d) again.

For Move I, apply Move II to the left side to get a Hopf link with both components labeled by \(\Omega\), one having a +1 framing. By Proposition (2b), \(F\) of this is \(q\text{dim}(\Omega)q\text{dim}^{-1/2}(\omega) = 1\).

b. This follows immediately from (a). \(\square\)

In particular, we have a symmetric bilinear pairing \(f_\varepsilon: V_g \otimes V_g \to \mathbb{C}\). If \(N_g = \{v \in V_g : f_\varepsilon(v, w) = 0 \ \forall w \in V_g\}\), define

\[
(14) \quad Z(\Sigma_g) = V_g/N_g,
\]

and then \(f_\varepsilon\) descends to a nondegenerate pairing on \(Z(\Sigma_g)\).
\[ \mathcal{F} \left( \begin{array}{c} \text{fragment} \\ \lambda_i \end{array} \right) = \sum_i N_{\lambda_i} \mathcal{F} \left( \begin{array}{c} \text{fragment} \\ \lambda_i \end{array} \right) / \text{qdim}(\lambda_i) = \sum_i N_{\lambda_i} \mathcal{F} \left( \begin{array}{c} \text{fragment} \\ \lambda_i \end{array} \right) = \mathcal{F} \left( \begin{array}{c} \text{fragment} \\ \lambda_i \end{array} \right) \]

**Figure 10. Proof of invariance under Move II**

**Proposition 5.** \( f_{[N,L]} : \otimes_j V_{g_j} \to \mathbb{C} \) descends to a map \( Z([N,L]) : \otimes_j Z(\Sigma_{g_j}) \to \mathbb{C} \).

**Sketch of Proof.** This follows from the fact that we can treat \( f_{[N,L]}(h_1 \otimes \cdots \otimes h_n) \) as \( f_{\varepsilon_{g_j}}(h_j \otimes k_j) \), where \( k_j \) is the image of the other \( h_i \)'s and \( L \) under an identification of the complement of the image of \( H_{g_j} \) with \( H_{g_j} \).

We are now closing in on our prey. The only question remaining is the elusive one of sewing and mending.

**Lemma 1.** For any sequence of signed, labeled points, there exists \( k \) and a set of intertwiners \( x_i \) and \( y_i \) such that the equalities hold in Figure 11 for the value of \( F \) on any closed, labeled, ribbon graph containing the pictures.

**Proof.** The sequence of signed labeled points corresponds to some tensor product of representations, say \( W \). Write the identity on \( W \) as a sum \( \sum_j p_j \) of projections onto irreducible subrepresentations (plus a projection onto a trivial trace subrepresentation which does not affect \( F \) and which we ignore). For the first, we may interpret the fragment \( T \) of the graph as a tangle, with \( F(T) \) intertwining \( W \) and the trivial representation. Thus \( F(T)p_i \) will be zero unless \( p_i \) projects onto a trivial subrepresentation. So we get \( F(T) = F(T) \sum_i p_i \), with the sum only over projections onto trivial subrepresentations. Writing each such \( p_i \) as \( y_i x_i \), with \( x_i \) an intertwiner from \( W \) to the trivial representation and \( y_i \) an intertwiner the other way, we have the result.

For the second, it follows from Proposition 2 that \( F \) of the tangle shown is a multiple of the projection onto the sum of all trivial subrepresentations. This equates the first and second pictures. The last is just Proposition 1e.

The following theorem first appears in [RT91], though not in the language of TQFTs. It was proven for \( sl_2 \), using only the Kauffman bracket formalism without quantum groups, in [BHMV95]. Complete proofs of the theorem as stated here first appear in [Wal] and [Tur94, Thm. IV.1.9].

**Figure 11. Cutting ribbon graphs**
Theorem 7. [RT91] Given a modular Hopf algebra $A$, the $Z$ defined in this section extends to a TQFT.

Proof. We need to check axioms (a)-(d) from Theorem 5. Nondegeneracy we have already checked, and symmetry is immediate from the construction.

For sewing, let $[M, L] = m$ and $[N, K] = n$ be presentations in standard form of the cobordisms mentioned in the sewing axiom. We may as well assume that each has only one boundary component, since we can check the equality with the maps applied to arbitrary vectors by gluing ribbon graphs into each other boundary component. Let $J$ be the presentation of their sewing described in Section 6, and let $j$ be the closed labeled ribbon graph obtained by labeling every component of $J$ by $\Omega$. Thus

$$Z(m \cup_s n) = K^{-1} F(j).$$

More precisely, since $[M, L]$ is in standard form, we can glue the complement of the image of $M$ to the underlying manifold of $dg$ and identify the result with $Hg$. This identification sends $L$ to a ribbon graph $h_m$ in $H_g$, such that $Z(m)(h) = \langle h_m, h \rangle$. Likewise we can define $h_n$ so that $Z(n)(h) = \langle h_n, h \rangle$. Now it is clear from the construction that $j$ is $h_m$ glued to $h_n$ via $dg$, so that

$$F(j) = K \langle h_m, h_n \rangle.$$

On the other hand

$$\langle h_m, h_n \rangle = \sum_{j=1}^{k} \langle h_m, a_j \rangle \langle b_j, h_n \rangle$$

$$= \sum_{j=1}^{k} Z(m)(a_j) Z(n)(b_j)$$

which gives the result.

For mending, let $[M, L] = m$ be a presentation in standard form of the cobordism mentioned in this axiom, and as above assume $m$ has only two boundary components, both of genus $g$. Let $a_j$ and $b_j$ be a basis and dual basis for $Z(\Sigma_g)$, and choose linear combinations of closed ribbon graphs in $H_g$ to represent them, which we will also call $a_j$ and $b_j$. The proof is then contained in Figure 12. The second equality is by Lemma 1a, the third by Proposition 1e and the seventh by Lemma 1b. The combinations $a_j$ and $b_j$ are each represented pictorially as one graph for conservation of subscripts: It does not affect the proof.

Remark 5. • These issues of biframing are easy to fix for closed manifolds. If $[\emptyset, L]$ is the manifold presented by $L$ and $\sigma(L)$ is the signature of the linking matrix of $L$, then a computation shows that $C^\sigma(L)Z([\emptyset, L])$ is invariant under framing if $C = 1 / u_+ (\Omega)$. This is the Reshetikhin-Turaev invariant, except they multiply by $K$ to get $F$ instead of $Z$, thus making the invariant multiplicative in connect-sums of manifolds. Their normalization is a bit different, because they in effect label by $\omega$ instead of $\Omega$ and thus must correct for the number of components as well as the signature of $L$.

• The vector spaces $Z(\Sigma_g)$ have been defined somewhat abstractly, and it is worth noting that they can be constructed quite explicitly. For example, it follows from what we’ve done that $Z(\Sigma_3)$ is spanned by graphs as shown in Figure 13, where $\lambda$, $\gamma$, and $\delta$ are any irreducible representations and $f_1, f_2$ are
any appropriate intertwiners. Thus $Z(\Sigma_2) \cong \bigoplus_{i,j,m=1}^k W^j_{i,i} \otimes W^j_{m,m}$, where $W^j_{i,j}$ is the space of intertwiners from $V_{\lambda_i} \otimes V_{\lambda_j}$ to $V_{\lambda_m}$, a space of dimension $N_{i,j}^m$.

By necessity the material presented here left off much of interest in the general field, and gave only a broad sketch of what it did cover. Below we offer the interested reader some pointers to other important topics and more details.

Good sources for classical knot theory include the excellent but dated [Rol76] and the more up-to-date [BZ85]. A good elementary survey of the knot polynomials can be found in [LM88].

Quite a lot is being done with quantum groups, with an eye towards both topological and algebraic applications, such as Kashiwara and Lusztig’s canonical bases and the Kazhdan-Lusztig conjectures. Fortunately, the field has recently benefited from a number of expository books, including [Lus93, CP94, Kas94, SS94], which cover all of this, and all of the background and details skipped in Sections 2 and 3. [Bax82] gives a good account of the integrable models in statistical mechanics, [Maj90, Fad84] connect them to quantum groups, and [Kau91] presents many ideas of statistical mechanics from a very knot-theoretic viewpoint.

Much has also been written about the 3-manifold invariants constructed in [RT91]. The $sl_2$ invariant has been constructed in many different forms and from many different perspectives in [KM91, Lic91, Cra91, BHMV92, BHMV95, Koh92].

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure12}
\caption{Proof of mending invariance}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure13}
\caption{Spanning labeled ribbon graphs in $H_2$}
\end{figure}
The square norm of the 3-manifold invariant associated to any modular Hopf algebra can be computed as a certain sum over states on a triangulation of the 3-manifold [TV92, Wal, Rob95]. There are also a number of 3-manifold invariants which are constructed from the same sorts of data but which appear to be different, including [Kup91, Kup, KR95].

The unfortunately unpublished [Wal] gives a thorough and informative account of the TQFTs we construct, extending the formalism to cobordisms, with corners (this is a strong version of the ‘duality’ which Witten uses to get his solution). It is a good resource for novices and experts. Much general information can be found in Quinn’s lecture notes on the subject [Qui95]. The other articles in the same volume represent an excellent introduction to many of the physical and geometric aspects of this subject.

There are other ways to get at the link and 3-manifold invariants quite apart from quantum groups. Affine Lie algebras [KP84, KW88], by way of conformal field theory [TK88, Ver88], offer a functor from the tangle category to a sort of intermediate linear category. One can still construct the TQFT in this context. This approach is technically more difficult, because there are infinite-dimensional algebras and vector spaces to contend with, but is directly linked to the physics.

Also closer to the physics are efforts to bring the tools mathematical physics has developed for understanding quantum field theories to bear on Chern-Simons theory. Good overviews of this include [Axe91, Ati90a]. More recent efforts have focused on perturbative approaches to Chern-Simons theory, including [AS92, AS94, Kon94]. Perturbation theory when applied to the link invariants is entirely combinatorial and fits naturally into the framework of Vassiliev invariants, which were developed out of purely topological considerations [Vas90, Vas92]. Good sources for the theory of Vassiliev invariants and their relation with the link invariants in this paper are [BN95, BL93], as well as the excellent expository article which recently appeared in these pages [Bir93] (it also gives a good introduction to the link invariants and their place in the history of knot theory). Recently there has been much activity around Vassiliev-type invariants of 3-manifolds, which appear to contain some perturbative information about the Chern-Simons invariants in purely combinatorial form [Oht95, Oht96, Roz, Gar, GL, Lin].

References


442 STEPHEN SAWIN


J. Roberts. The Kirby calculus for manifolds with boundary. preprint.


S. Sawin. Links, tangles and three manifolds: An invitation to quantum topology. lecture notes.


Department of Mathematics, Room 2-265, Massachusetts Institute of Technology, Cambridge, MA 02139-4307

E-mail address: sawin@math.mit.edu

Current address: Bannow 105, Department of Mathematics, Fairfield University, Fairfield, CT 06430-5195

E-mail address: ssawin@fairfield.edu