The Volume Conjecture

Sam Lewallen

July 5, 2008

Contents

1 Introduction

It is a fundamental goal of modern knot theory to "understand" the Jones polynomial. More specifically, following Jones's initial discovery in 1984, a plethora of knot and 3-manifold invariants were uncovered, known collectively as "quantum invariants." Furthermore, immediately upon the discovery of these invariants, it was recognized that they would have a dazzling variety of connections to diverse areas of mathematics and 2-dimensional physics. For example, they can all be described as representations of the braid group, followed by the application of a special trace; so the algebra containing the image of these representations would be important—examples include the Temperley-Lieb algebra, and the Birman-Wenzl-Murakami (BMW) algebra. Thus classical algebra is relevant. What's more, the images of generators in these representations are examples of R-matrices, which play an important role in solving statistical mechanical models and quantum integrable systems in 2 dimensions. So 2-dimensional physics is relevant. (Indeed, going even further with this point of view, one can take a diagram of a knot and actually define a statistical mechanical model *on the knot diagram* to get the same quantum invariants. Thus the combinatorics of knot diagrams are quite relevant.) To discover more R-matrices, Jimbo, Drinfel'd and others developed the formalism of quantum groups, specifically, quantum universal enveloping algebras—Hopf deformations $U_q(\mathfrak{g})$ of Lie algebras $\mathfrak{g}$ which carry R-matrices on their modules. This development really sparked the field, implying the existence of a vast number of distinct, computable knot invariants, indexed by (at least) the semi-simple finite-dimensional Lie algebras. The algebraic theory of quantum groups itself has become so vast that it is now rightly considered its own field; and for every development in that field, there is no reason to believe that there might not be some corresponding application to knot theory. More recently, the theory of Vassiliev or finite-type invariants, which studies an entire space of maps $S^1 \to S^3$ at once, has been shown to be intimately related with quantum invariants; here topology is relevant, but of infinite dimensional spaces.

What is decidedly *missing* from the above (incomplete) list is a connection with 3-dimensional topology itself. Indeed, even though one can get 3-manifold invariants from this framework (the so-called Witten-Reshitikhin-Turaev invariants, see [26], for instance), they are defined by surgery on a knot, and therefore share the short-comings of the quantum knot invariants: an explicit emphasis on algebra and 2 dimensions, rather than 3. Thus to "understand" the Jones polynomial is to identify its intrinsically 3-dimensional context.

In this paper, we present a fascinating conjecture in this direction, the *Volume Conjecture*, and explain how one might go about proving it. The Volume Conjecture was initially formulated by Rinat Kashaev in [8]. His ideas started along the previous lines: in fact, he used the quantum group perspective which we have mentioned, and it would have initially seemed that his invariant should fall to the way-side along with so many of the quantum invariants, which can be computed beautifully but give no readily apparent topological information. However, his invariant was explicitly constructed to quantize the dilogarithm function, which he knew was relevant to computing hyperbolic volume. This suggested to him that in the classical limit, his invariant should give the hyperbolic volume of the knot complement, and indeed, he computed this non-rigorously in several cases and confirmed it. What's more, Murakami and Murakami showed in [17] that Kashaev's invariant actually agrees with the colored Jones polynomial, which is the quantum invariant derived from $\mathfrak{s}l_2$, and parameterized by an integer N. Explicitly, calling this invariant F_N and letting M denote the complement of the knot in S^3, assumed to have a complete hyperbolic structure, the new form of the conjecture read

$$2\pi \times \lim_{N\to\infty} \frac{|F_N|}{N} = \mathrm{vol}(M)$$

Things had started to look interesting.

Our specific goal is to describe a framework in which the state-sum model for the colored Jones polynomial, calculated from a knot diagram, is shown to agree, asymptotically, with the hyperbolic volume of M, computed via a specific hyperbolic triangulation. It is obvious from the construction that the algebra of the quantum knot invariants contains deep information about the hyperbolic structure, an exciting realization; but the exact relationship remains unclear. It does not help things that the expositions of this approach to the conjecture to be found in the literature tend to be unclear and fragmented; furthermore, the necessary background is scattered across distinct and dense references, due to the conjecture's interdisciplinary nature. We hope that our exposition will make this fascinating conjecture and its ramifications more accessible.

In fact, even before Kashaev, it was already known that there was a 3-dimensional framework for the colored Jones polynomial, constructed by Turaev in [25], where one essentially takes a state-sum over a triangulation of the manifold M (this "essentially" turns out to be a rather thorny caveat, at least computationally). This should, in principle, given an even more direct framework for the Volume Conjecture; as of yet, this perspective has not been pursued successfully. If it were, it could be seen as a sort of simplicial approximation to Witten's famous Chern-Simons path integral definition of the Jones polynomial [28], which *is* 3-dimensional, and quite beautiful, but unfortunately is defined only at the physical level of rigor. Many generalizations of the Volume Conjecture reinforce this connection; other generalizations attribute additional hyperbolic structure to the quantum invariants, for example, determination of deformations of the complete structure. For more information about such generalizations, see [18, 30, 6, 5]. Furthermore, it is suggested in [17] that the proper generalization of the conjecture to non-hyperbolic knots is to replace $\mathrm{vol}(M)$ by the simplicial norm; if this generalization holds, it is an exciting corollary that the Vassiliev invariants determine the unknot.

In this paper, we concentrate on the background to the original conjecture, giving an informal overview of the machinery needed to compute the quantum knot invariants in the first place, and of the machinery needed to understand the hyperbolic volume computation. We choose to use the categorical framework of the Ribbon category, the Reshitikhin-Turaev functor, and quantum groups, necessitating, unfortunately, that we neglect many of the other frameworks which are available; it would be quite fascinating to understand how they themselves are related to the hyperbolic volume.

In §2, we give an overview of Hopf algebras and ribbon categories. Anyone familiar with these concepts can safely skip over it. In §3 we present the framework of the Reshitikhin-Turaev functor and the derivation of knot invariants from the ribbon category. In §4 we present the quantization of $\mathfrak{sl}_2$ and its ribbon Hopf algebra structure. We use this to produce the colored jones polynomial in §5, and describe how to compute it from a state-sum model. In §6 we give the necessary background in hyperbolic geometry. Throughout, we omit many proofs, placing emphasis on the ideas and their interrelationships. Finally, in §7, our true work begins: we present the Volume Conjecture, and the tantalizingly explicit relationship between the algebra of its state-sum formulation and the combinatorics of a hyperbolic triangulation of the knot complement.

The one thing we assume is the most rudimentary of knowledge about knot theory. This amounts essentially to familiarity with the Reidemeister moves, and what a skein relation is.

2 Ribbon Hopf Algebras and their Representations

We present the important algebraic properties of Hopf algebra representations, and abstract them to the setting of general categories. We follow [15] and [2].

Fix a field k. First, recall the definition of an algebra:

Definition 2.1. An *associative k-algebra A* with unit is a triple consisting of (A, m, η), where

1. A is a vector space over k.

2. m is a vector space map $m : A \otimes A \to A$, "multiplication," written $m(a \otimes b) = ab$, which is associative: $(ab)c = a(bc)$, or in other words, $m \circ (m \otimes \mathrm{id}) = m \circ (\mathrm{id} \otimes m)$ as maps from $A \otimes A \otimes A \to A$.

3. η is a vector space map $\eta : k \to A$ satisfying $a\eta(1) = a = \eta(1)a$. Taking $\lambda : k \otimes A \to A$ and $\rho : A \otimes k \to A$ to be the canonical isomorphisms, we can write this condition as $m \circ (\eta \otimes \mathrm{id}) \circ \lambda^{-1} = \mathrm{id} = m \circ (\mathrm{id} \otimes \eta) \circ \rho^{-1}$, as maps $A \to A$. Both η and $\eta(1)$ are called the *unit* of A, the latter written $1 \in A$

Remark 2.1. More generally, we could work with modules over a commutative ring, but we will stick with fields and vector spaces for convenience.

Flipping all the maps, we get the "dual" notion of a coalgebra:

Definition 2.2. A *coassociative k-coalgebra A* with counit is a triple consisting of (A, Δ, ϵ), where

1. A is a vector space over k

2. Δ is a vector space map $\Delta : A \to A \otimes A$, "comultiplication," which is coassociative: $(\mathrm{id} \otimes \Delta) \circ \Delta = (\Delta \otimes \mathrm{id}) \circ \Delta$ as maps $A \to A \otimes A \otimes A$

3. ϵ is a vector space map $\epsilon : A \to k$, "counit." Letting $\lambda : k \otimes A \to A$ and $\rho : A \otimes k \to A$ be the canonical isomorphisms as before, ϵ satisfies $\lambda \circ (\epsilon \otimes \mathrm{id}) \circ \Delta = \mathrm{id} = \rho \circ (\mathrm{id} \otimes \epsilon) \circ \Delta$ as maps $A \to A$.

We pause for a brief digression on notation: because $A \otimes A$ is spanned by the set $\{a \otimes a' : a, a' \in A\}$, we can write $\Delta(a) = \sum_{i=1}^{n} a_i \otimes a_i'$ in some (non unique) way. There is a short-hand for this, called Sweedler's notation, where we drop the indices and merely remember which factor is which: $\Delta(a) = \sum a_{(1)} \otimes a_{(2)}$. If desired, even the summation symbol can be dropped, $\Delta(a) = a_{(1)} \otimes a_{(2)}$, although we will not do this. When we compose comultiplications, we use multiple subscripts: for example, the value of $(\Delta \otimes \mathrm{id}) \circ \Delta$ on a is $\sum\sum a_{(1)(1)} \otimes a_{(1)(2)} \otimes a_{(2)}$, where we have written $a_{(1)(1)}$ for $(a_{(1)})_{(1)}$ and $a_{(1)(2)}$ for $(a_{(1)})_{(2)}$, and this generalizes to arbitrarily many subscripts. We can then express the coassociativity of Δ, and the counit property of ϵ, in Sweedler notation: the first becomes

$$\sum\sum a_{(1)(1)} \otimes a_{(1)(2)} \otimes a_{(2)} = \sum\sum a_{(1)} \otimes a_{(2)(1)} \otimes a_{(2)(2)} =: \sum a_{(1)} \otimes a_{(2)} \otimes a_{(3)},$$

and the second becomes

$$\sum \epsilon(a_{(1)})a_{(2)} = a = \sum a_{(1)}\epsilon(a_{(2)})$$

In this paper, we'll be interested in the case where the vector space A has an algebra *and* a "compatible" coalgebra structure. To make this compatibility precise, note that we can transfer any algebra structure from A to $A \otimes A$: if A is an algebra, we get a multiplication on $A \otimes A$ via

$$m_{A\otimes A}(a \otimes b, c \otimes d) = ac \otimes bd$$

for $a, b, c, d \in A$, and a unit via

$$\eta_{A\otimes A}(1) = 1 \otimes 1$$

for $1 \in A$. An analogous construction transfers any coalgebra structure from A to $A\otimes A$. Therefore, if A is both an algebra and coalgebra, so will $A \otimes A$ be, and we can require that Δ, ϵ be algebra maps, e.g.

$$\Delta(ab) = m_{A\otimes A}(\Delta(a), \Delta(b)) = m_{A\otimes A}\left(\sum a_{(1)} \otimes a_{(2)}, \sum b_{(1)} \otimes b_{(2)}\right) = \sum\sum a_{(1)}b_{(1)} \otimes a_{(2)}b_{(2)},$$

and that m, η be coalgebra maps. These structures are then said to be *compatible*, and A is called a *bialgebra*. Finally, we want one more piece of structure, a sort of "inverse" S, relating the multiplication and comultiplication.

Definition 2.3. A *Hopf algebra* is a bialgebra $(A, m, \eta, \Delta, \epsilon)$ along with a vector space map $S : A \to A$, the "antipode," satisfying

1. $\sum S(a_{(1)})a_{(2)} = \eta(\epsilon(a)) = \sum a_{(1)}S(a_{(2)})$. In other words, $\eta\circ\epsilon = m\circ(\mathrm{id}\otimes S)\circ\Delta = m\circ(S\otimes\mathrm{id})\circ\Delta$

Remark 2.2. Furthermore, it follows from the definition, after easy computations, that S is an antialgebra map, an anticoalgebra map, and is unique (uniqueness is proved in much the same way as the uniqueness of a group inverse).

Example 1. Suppose $\mathfrak{g}$ is a finite-dimensional Lie algebra over $\mathbb{C}$. We can then form an associative algebra with unit, the tensor algebra

$$T(\mathfrak{g}) = \mathbb{C} \oplus \mathfrak{g} \oplus (\mathfrak{g} \otimes \mathfrak{g}) \oplus (\mathfrak{g} \otimes \mathfrak{g} \otimes \mathfrak{g}) \oplus \cdots \tag{1}$$

with multiplication coming from the natural tensor product of two elements in $T(\mathfrak{g})$, and the unit being $1 \in \mathbb{C}$. There is a natural inclusion $\mathfrak{g} \subset T(\mathfrak{g})$ by sending $\mathfrak{g}$ to the second summand in (1); the elements in the image of this inclusion are called *primitive*. Evidently, the primitive elements generate $T(\mathfrak{g})$ as an algebra, and we write this multiplication in the standard way, using juxtaposition: e.g., the product of two primitive elements v, w is written vw.

Furthermore, we can put a Hopf algebra structure on $T(\mathfrak{g})$. Indeed, it is sufficient to define the comultiplication, counit, and antipode on primitive elements of $T(\mathfrak{g})$, and then extend them by linearity and as algebra maps in the first two cases, and by linearity and as an antialgebra map in the case of the antipode. Explicitly, letting $v \in T(\mathfrak{g})$ be a primitive element, we have

$$\Delta(v) = v \otimes 1 + 1 \otimes v,\ \ \epsilon(v) = 0,\ \ S(v) = -v$$

where of course the tensor product is the one from the expression $\Delta : T(\mathfrak{g}) \to T(\mathfrak{g}) \otimes T(\mathfrak{g})$. If one has never seen Hopf algebras before, it is a good exercise to check that this does in fact define a Hopf algebra structure.

The important thing about this Hopf algebra structure for us is that it descends to the universal enveloping algebra. Explicitly, if we let $I \subset T(\mathfrak{g})$ be the two-sided algebra ideal generated by the elements $vw - wv - [v, w]$, for all primitive elements $v, w \in \mathfrak{g} \subset T(\mathfrak{g})$ ($[,]$ is the Lie bracket), then one can check that, with the above Hopf algebra structure, I is in fact a Hopf algebra ideal[1]. Then the universal enveloping algebra $U(\mathfrak{g}) = T(\mathfrak{g})/I$ becomes a Hopf algebra, with comultiplication, counit, and antipode exactly the same as those for $T(\mathfrak{g})$.

Remark 2.3. As an irrelevant aside, note that there is an intrinsic way to define the Hopf algebra structure on $U(\mathfrak{g})$ using the universal perspective. Briefly, let ϕ be the canonical isomorphism from $U(\mathfrak{g} \oplus \mathfrak{g}) \to U(\mathfrak{g}) \oplus U(\mathfrak{g})$, let $\alpha : \mathfrak{g} \to \mathfrak{g} \oplus \mathfrak{g}$ be the diagonal map $v \mapsto (v, v)$, and let $U(\alpha)$ be the unique map $U(\mathfrak{g}) \to U(\mathfrak{g} \oplus \mathfrak{g})$ which exists by the universal property of $U(\mathfrak{g})$. Then one can check that $\Delta = \phi \circ U(\alpha)$. Similarly, with $\beta : \mathfrak{g} \to \{0\}$ being the unique map into the unique 0-dimensional Lie algebra, and $\gamma : \mathfrak{g} \to \mathfrak{g}^{\mathrm{op}}$ being the canonical isomorphism between a Lie algebra and its opposite, we have $\epsilon = U(\beta), S = U(\gamma)$.

The most important feature of $U(\mathfrak{g})$ turns out to be a defect, in a sense that we shall see in the course of this section and the next. Let $\tau : U(\mathfrak{g}) \otimes U(\mathfrak{g})$ be the flip map defined by $\tau(v \otimes w) = w \otimes v$. Then defining $\Delta^{op} = \tau \circ \Delta$, it can be easily seen that in the above example, $\Delta = \Delta^{op}$, i.e., $U(\mathfrak{g})$ is *cocommutative*. Furthermore $S^2 = \mathrm{id}$, and it can be shown that this is always a consequence of cocommutativity. A large part of the history of the study of Hopf algebras was spent looking for natural examples which are neither commutative *nor* cocommutative; in this paper, we will see that the existence of such algebras has deep ramifications in knot theory. By the end of §3, it will be evident that cocommutative Hopf algebras give no topological information, from our point of view.

We return now to general Hopf algebras; our interest in them stems from the excellent algebraic properties of their representations. First, let A be an algebra, and recall the definition of an A-module or representation of A:

Definition 2.4. Given a vector space V, a *representation* of A into V is an algebra map $\rho_V : A \to \mathrm{End}(V)$. V is called an A-module.

Given $a \in A$, we'll write the action of $\rho(a)$ on $v \in V$ as $a.v \in V$ for $v \in V, a \in A$. We also have morphisms between A-modules:

Definition 2.5. An *A-linear map* $f : V \to W$ between A-modules is a vector space map satisfying

$$a.f(v) = f(a.v)$$

i.e. a map which commutes with the actions of A on V and W.

Remark 2.4. One can similarly define right A-modules via an algebra map from $A^{op} \to \mathrm{End}(V)$, and there are also dual notions of left and right comodules for coalgebras.

With A-linear maps as the morphisms, the set $\mathbf{Rep}_A$ of all A-modules has the structure of a category, equipped with a natural forgetful functor to the category of vector spaces. Likewise, $\mathbf{Rep}_A^{fin}$ is the category of finite-dimensional A-modules.

[1] That is, $I \subset \ker(\epsilon), \Delta(I) \subset I \otimes T(\mathfrak{g}) + T(\mathfrak{g}) \otimes I$ (the appropriate duals to the conditions on an algebra ideal), and furthermore, $S(I) \subset I$.

When A is a Hopf algebra, its coalgebra and antipode give rise to extra structure on these categories, in the form of tensor products and duals. For any algebra A, if V and W are A-modules, then $V \otimes W$ is an $A \otimes A$ module via

$$(a \otimes a').(v \otimes w) = a.v \otimes a'.w$$

If A is a Hopf algebra, it is then straightforward to check that

$$a.(v \otimes w) := \Delta(a).(v \otimes w) \tag{2}$$

turns $V \otimes W$ into an A-module (use the compatibility of Δ and m). Likewise, the counit ϵ turns k into an A-module, the "trivial" A-module, via

$$a.1 := \epsilon(a) \tag{3}$$

for $1 \in k$ (and of course, in both cases, extended by linearity).

The importance of these "coalgebra" representations, in contrast to trivial representations on tensors and k such as $a.(v \otimes w) := a.v \otimes w$ and $a.1 := 1$, which exist for all algebras A, lies in the fact that by using the coalgebra structure, many canonical vector space maps become A-linear, so that they can descend to $\mathbf{Rep}_A$. For example, consider the canonical isomorphism $\lambda : k \otimes V \to V$, defined by $\lambda(1, v) = v$. The condition that λ be A-linear is

$$\lambda(a.(1 \otimes v)) = a.\lambda(1 \otimes v) \tag{4}$$

Given the module structure on tensor products from (2) and on k from (3), we transform the left hand side of (4) as follows:

$$\lambda(\Delta(a).(1 \otimes v)) = \lambda(a_{(1)}.1 \otimes a_{(2)}.v) = \lambda(\epsilon(a_{(1)}) \otimes a_{(2)}.v) = \epsilon(a_{(1)})(a_{(2)}.v) = \left(\epsilon(a_{(1)})(a_{(2)})\right).v$$

(with the last equality following from the multiplicativity property of the representation). The right-hand side of (4) is simply $a.v$. In other words, (4) gives us precisely the counit axiom $a = \epsilon(a_{(1)})a_{(2)}$.

Thus, when A is a bialgebra, there is an entire tensor product structure on $\mathbf{Rep}_A$ and $\mathbf{Rep}_A^{fin}$ (we have not yet used the antipode of a Hopf algebra). We want to abstract this for general categories:

Definition 2.6. A *monoidal category* is a category $\mathcal{C}$ equipped with

1. a functor $\otimes : \mathcal{C} \times \mathcal{C} \to \mathcal{C}$ called the tensor product, with $\otimes(V, W)$ written $V \otimes W$
2. an object I called the unit object
3. a natural isomorphism, "associativity," written α, between functors $\otimes \circ (\mathrm{id} \times \otimes) : \mathcal{C} \times \mathcal{C} \times \mathcal{C} \to \mathcal{C}$ and $\otimes \circ (\otimes \times \mathrm{id}) : \mathcal{C} \times \mathcal{C} \times \mathcal{C} \to \mathcal{C}$. Its components $\alpha_{A,B,C}$ must satisfy the "coherence property"

$$\begin{tikzcd}
((A \otimes B) \otimes C) \otimes D \arrow[r, "\alpha_{A,B,C\otimes D}"] \arrow[d, "\alpha_{A\otimes B,C,D}"'] & (A \otimes (B \otimes C)) \otimes D \arrow[r, "\alpha_{A,B\otimes C,D}"] & A \otimes ((B \otimes C) \otimes D) \arrow[d, "A\otimes\alpha_{B,C,D}"] \\
(A \otimes B) \otimes (C \otimes D) \arrow[rr, "\alpha_{A,B,C\otimes D}"'] & & A \otimes (B \otimes (C \otimes D))
\end{tikzcd}$$

where the same symbol refers either to an object of $\mathcal{C}$ or its identity morphism

4. natural isomorphisms, "left," respectively "right," identity, written λ and ρ, between functors $(I \otimes \cdot) : \mathcal{C} \to \mathcal{C}$ and $\mathrm{id} : \mathcal{C} \to \mathcal{C}$, and $(\cdot \otimes I) : \mathcal{C} \to \mathcal{C}$ and $\mathrm{id} : \mathcal{C} \to \mathcal{C}$, respectively. They must satisfy the coherence property

$$\begin{tikzcd} m\,(A \otimes I) \otimes B \arrow[rr, "\alpha_{A,I,B}"] \arrow[dr, "\rho_A \otimes B"'] & & A \otimes (I \otimes B) \arrow[dl, "A \otimes \lambda_B"] \\ & A \otimes B & \end{tikzcd}$$

The motivation for these particular coherence properties lies in a theorem of MacLane, which states that they imply the commutativity of all other "reasonable" diagrams. (See [14] for an introduction to this sort of category theory). Furthermore, if the isomorphisms α, λ, ρ are all identity morphisms, then we call $\mathcal{C}$ a *strict* monoidal category, and another important theorem of MacLane states that every monoidal category is equivalent to a strict one. Thus, if we only care about our categories up to equivalence, there is no loss of generality in assuming they are strict. This saves much suppression-of-isomorphisms (or even worse, much actually-writing-out-of-isomorphisms) when writing out identities in monoidal categories.

Now, our previous discussion on the representation theory of bialgebras extends to

Theorem 2.1. *If A is a bialgebra, then $\mathbf{Rep}_A$ and $\mathbf{Rep}_A^{fin}$ are monoidal categories via* (2) *and* (3)*, k being the identity object.*

The necessary coherence conditions follow from coassociativity of Δ and the counit axioms, and are easily, albeit tediously, checked.

After tensor products, we would like duals. To avoid issues with dual bases, we will restrict our attention to finite-dimensional vector spaces and $\mathbf{Rep}_A^{fin}$; at first, let us consider $\mathbf{Rep}_A^{fin}$ composed with its forgetful functor to the category of vector spaces. Every object $V \in \mathbf{Rep}_A^{fin}$ has a dual vector space $V^* = \mathrm{Hom}(V, k)$. For $v^* \in V^*, v \in V$, the dual pairing $\langle v^*, v \rangle = v^*(v)$ defines a linear map

$$\mathrm{ev}_V : V^* \otimes V \to k, \qquad \mathrm{ev}(v^* \otimes v) = \langle v^*, v \rangle \tag{5}$$

and extended to $V^* \otimes V$ by linearity. Furthermore, in the finite dimensional case, if $\{e_i\}$ is a basis for V, then there is a unique dual basis $\{e^i\}$ defined by the pairing $\langle e^i, e_j \rangle = \delta_{ij}$. In this case we get a "copairing," a (linear) map

$$\mathrm{coev}_V : k \to V \otimes V^*, \qquad 1 \mapsto \sum_i e_i \otimes e^i \tag{6}$$

which, one should check, is actually independent of the choice of basis. These maps allow us to formally construct many of the interesting objects of linear algebra. For example, for each $f \in \mathrm{End}(V)$, we know there is a dual map $f^* \in \mathrm{End}(V^*)$ defined by $\langle f^*(v^*), v \rangle = \langle v^*, f(v) \rangle$. Indeed, using the above structure maps, this can be written

$$f^* : V^* = V^* \otimes 1 \overset{\mathrm{id} \otimes \mathrm{coev}}{\longrightarrow} V^* \otimes V \otimes V^* \overset{\mathrm{id} \otimes f \otimes \mathrm{id}}{\longrightarrow} V^* \otimes V \otimes V^* \overset{\mathrm{ev} \otimes \mathrm{id}}{\longrightarrow} 1 \otimes V^* = V^* \tag{7}$$

Writing out the above on a basis for V, it's an easy (and actually, fun) computation to see that the two definitions agree. (We'll go through the explicit details of a similar, but slightly more sophisticated, computation later on).

Remark 2.5. One might wonder how we chose the order of V and V^* in the tensor products in (5) and (6). Indeed, the choice we made is called a *left* dual. The choice was arbitrary, but what is important is that the two choices are not trivially the same, in general; it only seems that way because we have a flip map τ between tensor products in a vector space. We'll get to this point in a moment.

Let's return to A-modules: can the above structure be extended to $\mathbf{Rep}_A^{fin}$? Suppose the representation $\rho_V : A \to \mathrm{End}(V)$ defines an A-module structure on V; then, for each $a \in A$, the map ρ_V^*, defined by sending a to the dual map $(\rho_V(a))^* \in \mathrm{End}(V^*)$ of $\rho_V(a)$, is a natural candidate for an A-module structure on V^*. However, it is easy to check that this is a *right*-module, not left, due to the equality $(f \circ g)^* = g^* \circ f^*$. We need to compose with an antialgebra map to flip the factors, and the antipode saves the day: one can check that the map

$$\rho_{V^*}(a) = \rho_V(S(a))^*$$

is a well-defined representation. In terms of the dual pairing, we have

$$\langle a.v^*, v\rangle = \langle v^*, S(a).v\rangle$$

Just as in the tensor case, using this novel structure gives us more than we bargained for: one can check that the antipode axiom is precisely what's needed to ensure that ev_V and coev_V are A-linear for all modules $V \in \mathbf{Rep}_A^{fin}$ and the trivial module structure on k.

Just as for tensors, we can abstract this dual structure:

Definition 2.7. If $\mathcal{C}$ is a monoidal category and V an object of $\mathcal{C}$, an object V^* of $\mathcal{C}$ is said to be a *left dual* for V if there are morphisms $\mathrm{ev}_V : V^* \otimes V \to I$ and $\mathrm{coev}_V : I \to V \otimes V^*$ making the following diagrams commute:

$$\begin{tikzcd} V \arrow[r, "\mathrm{coev}\otimes\mathrm{id}"] & (V \otimes V^*) \otimes V \\ V \arrow[u, "\mathrm{id}"] & V \otimes (V^* \otimes V) \arrow[l, "\mathrm{id}\otimes\mathrm{ev}"'] \arrow[u, "\alpha"'] \end{tikzcd} \qquad \begin{tikzcd} V^* \arrow[r, "\mathrm{id}\otimes\mathrm{coev}"] \arrow[d, "\mathrm{id}"] & V^* \otimes (V \otimes V^*) \arrow[d, "\alpha"'] \\ V^* & (V^* \otimes V) \otimes V^* \arrow[l, "\mathrm{ev}\otimes\mathrm{id}"'] \end{tikzcd}$$

where we've suppressed certain isomorphisms (or not, in the strict case).

Suppose V, W are objects in $\mathcal{C}$ with left duals V^*, W^*, and $\psi : U \to V$ is a morphism of $\mathcal{C}$. Then the composite

$$V^* = V^* \otimes 1 \overset{\mathrm{id}\otimes\mathrm{coev}}{\longrightarrow} V^* \otimes U \otimes U^* \overset{\mathrm{id}\otimes\psi\otimes\mathrm{id}}{\longrightarrow} V^* \otimes V \otimes U^* \overset{\mathrm{ev}\otimes\mathrm{id}}{\longrightarrow} 1 \otimes U^* = U^*$$

defines a *dual morphism* $\psi^* : V^* \to U^*$ (this generalizes (7)). If we choose a unique left dual structure $V^*, \mathrm{ev}_V, \mathrm{coev}_V$ for every object V of some category $\mathcal{C}$, we obtain a contravariant functor $* : \mathcal{C} \to \mathcal{C}$ called the *dual object functor*, sending $V \mapsto V^*$, $\psi \mapsto \psi^*$.

Definition 2.8. A monoidal category $\mathcal{C}$ is said to be *rigid* if every object has a (uniquely made) choice of left dual, and if the dual object functor is an anti-equivalence of categories.

Remark 2.6. If $\mathcal{C}$ is a rigid monoidal category, then the anti-equivalence of $*$ implies the existence of a bijection from $\mathrm{Hom}(U \otimes V, W)$ to $\mathrm{Hom}(U, W \otimes V^*)$, ϕ being sent to the map

$$U = U \otimes 1 \overset{\mathrm{id}_U \otimes \mathrm{coev}_V}{\longrightarrow} U \otimes V \otimes V^* \overset{\phi \otimes \mathrm{id}_{V^*}}{\longrightarrow} W \otimes V^*$$

In particular, there is a bijection between morphisms $U \otimes V \to I$ and $U \to V^*$, and between morphisms $V \to W$ and $I \to W \otimes V^*$

Remark 2.7. One can check that there are natural isomorphisms between $(U \otimes V)^*$ and $V^* \otimes U^*$ for all $U, V \in \mathcal{C}$.

Remark 2.8. Note, however, that U^{**} is *not* isomorphic to U, in general.

Again, using the structure from (5) and (6), it straightforward to prove that

Theorem 2.2. *If A is a Hopf algebra, the category $\mathbf{Rep}_A^{fin}$ is rigid monoidal.*

Continuing on our quest to do all of linear algebra in our category $\mathbf{Rep}_A^{fin}$, we want to abstract the *trace* of a map $f : V \to V$. Again, let $\{e_i\}$ be a basis for V; then, the matrix-definition of trace is easily seen to be equivalent to $\mathrm{tr}(f) = \langle e^i, f(e_i) \rangle$, once we identify $V \cong V^*$ appropriately. Indeed, even in cases where no such identification exists, this previous definition gives rise to the more general

$$\mathrm{tr} = \mathrm{ev}_V \circ \tau_{V,V^*} \circ (f \otimes \mathrm{id}_{V^*}) \circ \mathrm{coev}_V : k \to k \tag{8}$$

where τ_{V,V^*} is the map $\tau_{V,V^*}(v \otimes v^*) = v^* \otimes v$. If we rewrite the original dual-pairing definition of trace, in the case of vector spaces, as a map $k \to k$ defined to take the value $\mathrm{tr}(f)$ on 1, it is another little exercise to see that it agrees with (8); one need only recall the definitions of coev and ev.

The formula from (8) almost gives an A-linear trace, except that the flip map $\tau_{V,W}$ is usually not A-linear (though it will be if A is cocommutative). What we need is for $\mathbf{Rep}_A^{fin}$ to have some *A-linear* flip, called a *braiding*:

Definition 2.9. A *braided* monoidal category $\mathcal{C}$ is a monoidal category with a natural isomorphism σ, the "braiding," between the functors $\otimes : \mathcal{C} \times \mathcal{C} \to \mathcal{C}$ and $\otimes \circ \tau : \mathcal{C} \times \mathcal{C} \to \mathcal{C}$ (τ being the flip functor) and a natural isomorphism α satisfying the coherence property

$$\begin{tikzcd}
 & A \otimes (B \otimes C) \arrow[r, "\sigma"] & (B \otimes C) \otimes A \arrow[dr, "\alpha"] & \\
(A \otimes B) \otimes C \arrow[ur, "\alpha"] \arrow[dr, "\sigma \otimes 1"'] & & & B \otimes (C \otimes A) \\
 & (B \otimes A) \otimes C \arrow[r, "\alpha"'] & B \otimes (A \otimes C) \arrow[ur, "1 \otimes \sigma"'] &
\end{tikzcd}$$

In general, the category $\mathbf{Rep}_A^{fin}$ need not be braided; indeed, the multiplications Δ (defining the representation on $V \otimes W$) and Δ^{op} (defining the representation on $W \otimes V$) can be essentially unrelated. On the other hand, when A is cocommutative, so that $\Delta = \Delta^{op}$, the standard vector space flip τ gives a braiding. In between these two cases is what is called a *quasitriangular* Hopf algebra, where Δ and Δ^{op} differ by a sort of inner automorphism:

Definition 2.10. A Hopf algebra A is *quasitriangular* if it is equipped with an invertible element $\mathcal{R} \in A \otimes A$ satisfying

1. $\mathcal{R}\Delta(a) = (\tau \circ \Delta)(a)\mathcal{R}$
2. $(\Delta \otimes 1)(\mathcal{R}) = \mathcal{R}_{13}\mathcal{R}_{23}$
3. $(1 \otimes \Delta)(\mathcal{R}) = \mathcal{R}_{13}\mathcal{R}_{12}$

where, writing $\mathcal{R} = \sum r_{(1)} \otimes r_{(2)}$ (Sweedler's notation), we have, for example, $\mathcal{R}_{12} = \sum r_{(1)} \otimes r_{(2)} \otimes 1 \in A \otimes A \otimes A$

$\mathcal{R}$ is called the *universal R-matrix.*

Theorem 2.3. *If A is a quasitriangular Hopf algebra, then $\mathbf{Rep}_A^{fin}$ is braided, with $\sigma_{V,W} : V \otimes W \to W \otimes V$ equal to $\tau_{V,W} \circ \rho_{V \otimes W}(\mathcal{R})$ where τ is the standard vector space flip, $\mathcal{R} \in A \otimes A$ is the universal R-matrix, and $\rho_{V \otimes W}$ is the natural representation of $A \otimes A$ on $V \otimes W$.*

The braiding property follows from 1. of Definition 2.10, and the coherence rules follow from 2. and 3.

Therefore, if A is a quasitriangular Hopf algebra, and $f : V \to V$ is a morphism of $\mathbf{Rep}_A^{fin}$, we define its *A-trace* to be the map

$$\mathrm{ev}_V \circ \tau_{V,V^*} \circ \rho_{V \otimes V^*}(\mathcal{R}) \circ \mathrm{coev}_V : k \to k$$

First we can ask: does this correspond to the trace of an operator, in the traditional matrix sense? Indeed, the answer is yes. Define $u \in A$ to be the element $m((S \otimes \mathrm{id})(\tau(\mathcal{R})))$, in other words, if $\mathcal{R} = \sum r_{(1)} \otimes r_{(2)}$, then

$$u = S(r_{(2)})r_{(1)} \tag{9}$$

Theorem 2.4. *If ρ_V is a finite-dimensional representation of a quasitriangular Hopf algebra A, and $f : V \to V$ is some A-linear map, then the A-trace of f evaluated at $1 \in k$ is equal to the standard matrix trace $tr(\rho_V(u) \circ f)$, where u is as in* (9).

Proof. We will compute each in terms of a basis $\{e_i\}$ for V. By the definition of matrix trace, we have

$$\mathrm{tr}(\rho_V(u) \circ f) = \langle e^i, u.f(e_i)\rangle = \langle e^i, (S(r_{(2)})r_{(1)}).f(e_i)\rangle \tag{10}$$

By examining its definition above, we see that the A-trace sends

$$1 \mapsto e_i \otimes e^i \mapsto r_{(1)}.e_i \otimes r_{(2)}.e^i \mapsto r_{(2)}.e^i \otimes r_{(1)}.e_i \mapsto \langle r_{(2)}.e^i, r_{(1)}.e_i\rangle$$

By the definition of the dual module, and then by multiplicativity of representations,

$$\langle r_{(2)}.e^i, r_{(1)}.e_i\rangle = \langle e^i, S(r_{(2)}).(r_{(1)}.e_i)\rangle = \langle e^i, (S(r_{(2)})r_{(1)}).f(e_i)\rangle$$

which agrees with (10), as desired. $\square$

One would also like the A-trace to respect tensor products, so that the trace of a tensor product of maps is the product of the traces of its summands. However, one can compute

$$\Delta(u) = (\tau(\mathcal{R})\mathcal{R})^{-1}(u \otimes u). \tag{11}$$

For the trace to be multiplicative, after going through the computations, one sees that we need $\Delta(u) = u \otimes u$.

Remark 2.9. This element u is important for another reason, namely because one can prove that it implements the antipode squared, by conjugation:

$$S^2(a) = uau^{-1} \tag{12}$$

for all $a \in A$. Now, note that there is a canonical isomorphism of vector spaces, $\phi_V : V \to V^{**}$, defined by $\langle \phi_V(e_i), e^j \rangle = \langle e^j, e_i \rangle$. When one checks whether this map is A-linear, one runs into the problem that $S^2(a) \neq a$. However, if we make a new map $\phi'_V = \phi_V \circ \rho_V(u)$, then we are instead hoping for the identity $S^2(a)u = ua$, which is equivalent to (12). Thus, for any quasitriangular algebra A, there are canonical isomorphisms between double duals in $\mathbf{Rep}_A^{fin}$. However, these isomorphisms ϕ' are not tensorial: we do not have $\phi'_{V\otimes W} = \phi'_V \otimes \phi'_W$, because $\Delta(u) \neq u \otimes u$, the same obstruction to multiplicativity of the A-trace.

Remark 2.10. There is yet another perspective on this whole story. Recall that in any rigid category, we have a left dual $\mathrm{ev}_V : V^* \otimes V \to I$. However, now that we have a twist $\sigma_{V^*,V}$, we have a candidate,

$$\widetilde{\mathrm{ev}}_V = \mathrm{ev}_V \circ \sigma_{V,V^*} : V \otimes V^* \to I \tag{13}$$

for a "right" dual. Moreover, by Remark 2.6, there is a corresponding morphism $J_V : V^{**} \to V$. One can check, from the details of the construction of Remark 2.6, that this map is exactly ϕ'_V (from Remark 2.9) when we're in $\mathbf{Rep}_A^{fin}$. Furthermore, this telling of the story was completely general, depending only on the structure of a rigid, braided category, and so we conclude that any such category $\mathcal{C}$ has isomorphic double duals via J_V. We still, however, have $J_U \otimes J_V = \sigma_{W,V} \circ \sigma_{V,W} \circ J_{V\otimes W}$ (instead of tensoriality—compare to (11)), and, likewise, the maps $\widetilde{\mathrm{ev}}_V$ are not tensorial and therefore do not define a right dual structure on $\mathcal{C}$. Finally, $\mathcal{C}$ also has an "A-trace,"

$$\widetilde{\mathrm{ev}}_V \circ (f \otimes \mathrm{id}_{V^*}) \circ \mathrm{coev}_V : k \to k$$

for all morphisms $f : V \to V$, which still does not respect tensor products.

The solution to our tensoriality troubles is to balance our category:

Definition 2.11. A rigid, braided category is called *balanced*, or *ribbon*, if it is equipped with an automorphism b_V of V, for every object $V \in \mathcal{C}$, such that

1. $b_U \otimes b_V = \sigma_{W,V} \circ \sigma_{V,W} \circ b_{V\otimes W}$
2. $b_{V^*} = (b_V)^*$
3. $b_I = \mathrm{id}_I$

Then we can check that the map $k_V = J_V \circ b_V^{-1}$ is an isomorphism $V \to V^{**}$ satisfying

$$k_{U\otimes V} = k_U \otimes k_V$$

(up to the isomorphism between the ranges, $(V \otimes W)^{**}$ and $V^{**} \otimes W^{**}$, of the left and right hand sides, given by Remark 6.). Likewise, the map

$$\widetilde{\mathrm{ev}}_V = \mathrm{ev}_V \circ (b_{V^*}^{-1} \otimes \mathrm{id}) \circ \sigma_{V,V^*} : V \otimes V^* \to I$$

gives a right dual structure, along with the similarly defined map $\widetilde{\mathrm{coev}}_V$ (we haven't given a definition of right dual structure, but it should be easy to deduce from Definitions 2.7 and **??**). Finally, for a map $f : V \to V$, the map

$$\mathrm{qtr}(f) = \widetilde{\mathrm{ev}}_V \circ (f \otimes \mathrm{id}) \circ \mathrm{coev}_V$$

replaces the old "trace" and is called the *quantum trace* of f; it is an easy exercise to check that it *is* tensorial. We define the *quantum dimension* of an object $V \in \mathcal{C}$ to be $\mathrm{qtr}(\mathrm{id}_V)$ $(= \widetilde{\mathrm{ev}}_V \circ \mathrm{coev}_V)$.

Definition 2.12. A *ribbon Hopf algebra A* is a quasitriangular Hopf algebra equipped with an invertible central element ν, the "ribbon" element, satisfying:

0. $\nu^2 = uS(u)$
1. $\Delta(\nu) = ((\tau(\mathcal{R})\mathcal{R})^{-1}(\nu \otimes \nu)$
2. $S(\nu) = \nu$
3. $\epsilon(\nu) = 1$

where u is as in (9), and the numbering is arranged to demonstrate the analogy with Definition 2.11.

Thus we see that in general, $\mathbf{Rep}_A^{fin}$ fails to be balanced because the element $uS(u)$ of A has no square root. If it does we call A ribbon, and the following should be no surprise:

Theorem 2.5. *If A is a ribbon Hopf algebra, then $\mathbf{Rep}_A^{fin}$ is a ribbon category with $b_U = \rho_U(\nu)$ (it is A-linear because ν is central), and the map k_V is $\phi_V \circ \rho_V(\nu^{-1}u)$, where ϕ_V is still the canonical vector space isomorphism $V \to V^{**}$.*

It is worth giving the element $\nu^{-1}u$ its own notation, though there is nothing standard in the literature. Let us write $\kappa = \nu^{-1}u$.

Remark 2.11. Just as in Theorem 2.4, we can prove that

$$\mathrm{qtr}(f) = \mathrm{tr}(\rho_V(\kappa) \circ f),$$

where $f : V \to V$ is a morphism from $\mathbf{Rep}_A^{fin}$ and A is a Ribbon Hopf algebra. Similarly, we can compute

$$\widetilde{\mathrm{ev}}_V(e_i \otimes e^j) = \left\langle e^j, \rho_V(\kappa).e_i \right\rangle$$

$$\widetilde{\mathrm{coev}}_V(1) = \sum_i e^i \otimes (\rho_V(\kappa))^{-1}.e_i$$

Remark 2.12. This close correspondence between categories of Hopf algebra modules and rigid, monoidal categories is an example of a general "Tannaka-Krein type" duality. Indeed, under certain conditions it can be shown that every rigid monoidal categories can be written as the category of representations for some Hopf algebra (see [2] for details).

Remark 2.13. Note that when we say a category $\mathcal{C}$ is ribbon, we mean implicitly that it is braided, rigid, and monoidal as well. Furthermore, we will refer to the defining morphisms of a ribbon category (the braidings, ribbons, eval and coeval maps, etc.) collectively as *structure maps*.

Finally, before we turn to some topology, let us try to relate these past results to our first example. However, there is not much to say: it is obvious that taking $\mathcal{R} = 1 \otimes 1$ makes any cocommutative Hopf algebra quasitriangular, and, because $S^2 = \mathrm{id}$ in this case, $\nu = 1$ is a "ribbon" element. Therefore $U(\mathfrak{g})$ is trivially a ribbon Hopf algebra, and we are left searching for non-trivial examples.

3 "Ribbon Category" and the Reshetikhin-Turaev Functor

There is a remarkable correspondence, relating tangles in $\mathbb{R}^3$ and the categorical machinery of the previous section, which gives both an intuitive understanding of many of the maps we went through defining, and a powerful framework for knot and tangle invariants. This material can be found in [2, 10]. The original paper reference is [21]. Diagrams are modifications of those in [2].

First, we place the set of tangles into a categorical setting. We assume the reader is familiar with the standard definition of tangles in $\mathbb{R}^2 \times [0, 1]$, and more precisely, (n, m)-tangles, with n strands at the top ($\mathbb{R}^2 \times \{0\}$) and m at the bottom ($\mathbb{R}^2 \times \{1\}$), defined up to ambient isotopy fixing the top and bottom planes. Furthermore, we will consider framed tangles, which we choose to think of as embeddings of rectangles and annuli, replacing line segments and circles (equivalently, a framed tangle is a smoothly embedded tangle along with a choice of framing of the normal bundle of the embedding). We will require these ribbon tangles to be orientable, which for us means there can be annuli but no Möbius bands, and the each side of each ribbon must face in the same direction at both top and bottom. We will also orient our tangles in the other sense, i.e., we make a choice of direction along the core of each ribbon, either towards the top or the bottom. (It is this meaning that we will be referring to when we mention orientation, for the rest of this section). One figure should suffice to make this all clear (in our depictions of tangles, we will assume 0 is at the top and 1 is at the bottom):

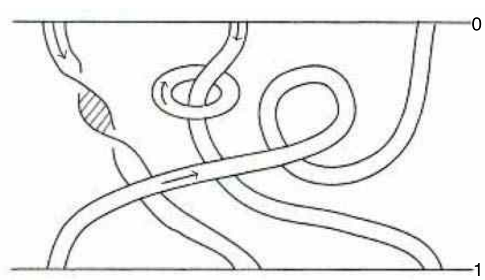

Figure 1: Diagram of a (3,3) oriented ribbon tangle, or "ribbon"

Such an oriented ribbon tangle R, which we will refer to as a *ribbon* from now on, intersects $\mathbb{R}^2 \times \{x\}$, for $x \in \{0,1\}$, in some number of ordered line segments, each of which inherits one of two orientations corresponding to the local orientation of R at that segment. Refer to the orientation agreeing with the direction from 0 to 1 as $+$ (the "downward" direction, in our figures), and the other direction as $-$ (upward). Then, by writing down the directions of the segments in order, we get two words on the symbols $+$ and $-$, call them w_R^{in} and w_R^{out}, corresponding to $x = 0$ and $x = 1$, respectively. For example, letting R be the tangle from Figure 1, we have

$$w_R^{in} = ++-, \quad w_R^{out} = -++$$

If we have two ribbons R and S such that $w_R^{out} = w_S^{in}$, then we get another ribbon, $S \circ R$, by stacking R on top of S, gluing them along the common plane, and then rescaling. Formally, let $\mathcal{O}$ be the set of all words on $\{+,-\}$, including the empty word $\emptyset$. Let **Rib** denote the set of (isotopy classes of) ribbons. Then we get a category, with objects $\mathcal{O}$, and with $\mathrm{Hom}(x,y) = \{R \in \mathbf{Rib} | w_R^{in} = x, w_R^{out} = y\}$ for $x, y \in \mathcal{O}$, with composition as just defined (the identity morphism on a word x is the ribbon consisting of vertical strips of the appropriate number and orientations). We will then abuse notation as follows:

Definition 3.1. **Rib** is the category just defined, with objects $\mathcal{O}$ and morphisms (also) **Rib**.

Remark 3.1. $\mathrm{End}(\emptyset)$ is precisely the set of isotopy classes of framed, oriented links.

Furthermore, we can define a tensor product on **Rib** as follows: the tensor of two sequences is just their concatenation, $w \otimes v = wv$, and the tensor of two tangles is just side-by-side juxtaposition (again, in our depictions, this will be left-to-right). It is then immediate that **Rib** forms a *strict* monoidal category, with identity object the empty sequence $\emptyset$. What's more, **Rib** is a ribbon category: taking the dual of a word reverses it and switches $+$'s and $-$'s; as for the rest, we've shown the structure morphisms for $+$ in Figure 2 (notation from §2):

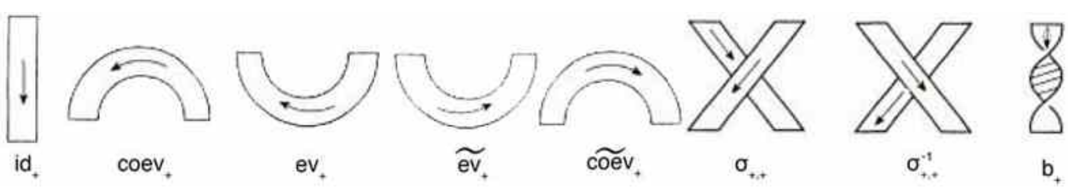

Figure 2: Structure morphisms for $+$

Taking the dual of a morphism rotates it by 180°; taking the dual of the above maps gives us the structure morphisms for $-$. The structure morphisms for other objects are determined uniquely by tensoriality considerations. It is not difficult to check that the defining relations for a ribbon, braided, rigid category are satisfied. For the proof, see [10].

Remark 3.2. This elucidates, for example, the picturesque terminology for braids and ribbons.

In fact, **Rib** can be considered *the* "free" (strict) ribbon category. To explain what we mean by this, we first note that every isotopy class of ribbons can be represented (highly non-uniquely) by a tangle diagram (as we have been doing); we shall think of these simply as regular projections onto the plane $\mathbb{R} \times \{0\} \times [0,1] \subset \mathbb{R}^2 \times [0,1]$. Let $\mathcal{T}$ denote the set of all the tangle diagrams in Figure 2, as well as all other tangle diagrams derived from those in Figure 2 by changing orientation on one or more of the strands. This set $\mathcal{T}$ is known as the set of *elementary tangles*. Then it is well-known (and not difficult to show, see [10] or [26]) that all ribbon diagrams can be factored into diagrams from $\mathcal{T}$ using the tensor product and composition operations that we have described. What's more, two such diagrams represent isotopic ribbons if and only if they are related by moves from the following list:

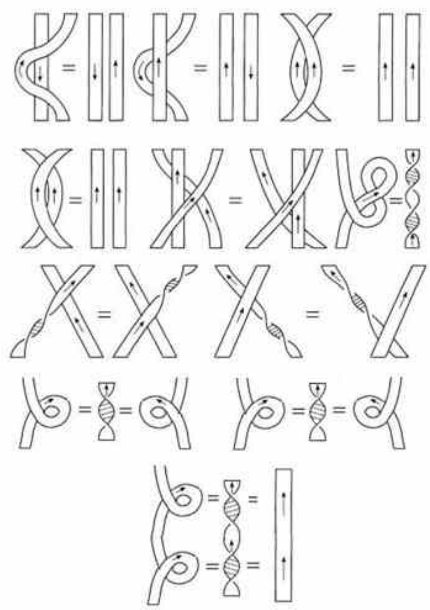

Figure 3: Relations between tangle generators

This is essentially due to Reidemeister; find proofs in the previous references, and references therein. Now, suppose we have a strict, rigid, monoidal category $\mathcal{C}$, with a choice of object $V \in \mathcal{C}$. To every word $w \in \mathcal{O}$ we associate an object in $\mathcal{C}$ by replacing each $+$ of w with V, each $-$ with V^*, and taking their tensor product in the corresponding order; denote this element as $\mathcal{F}_V(v)$ (to $\emptyset$ we associate the tensor identity object of $\mathcal{C}$). Suppose we also assign, to each of the morphisms $R : v \to w$ from $\mathcal{T}$, a morphism $\mathcal{F}_V(R) \in \mathrm{Hom}(\mathcal{F}_V(v), \mathcal{F}_V(w))$ of $\mathcal{C}$. Then the above discussion amounts to saying that this map $\mathcal{F}_V$ extends to a unique functor $\mathcal{F}_V : \mathbf{Rib} \to \mathcal{C}$, respecting tensor products, if and only if it sends each side of the relations from Figure 3 to the same morphism. Thus the following theorem, whose proof is technical and not difficult, and can be found in [10], justifies the statement that **Rib** is the free ribbon category:

Theorem 3.1. *Let $\mathcal{C}$ be a strict ribbon category, fix an object $V \in \mathcal{C}$, and let $\mathcal{F}_V$ be as defined above. Further, each element in $\mathcal{T}$ is a structure map in the obvious way, so define $\mathcal{F}_V$ on $\mathcal{T}$ by sending each map to the corresponding structure map in $\mathcal{C}$, e.g. $\mathcal{F}(ev_+) = ev_V, \mathcal{F}(\sigma_{+,+}) = \sigma_{V,V}$, etc. (and likewise for $-$). Then $\mathcal{F}_V$ extends to a functor $\mathbf{Rib} \to \mathcal{C}$, i.e., the relations are satisfied.*

Proof. It suffices to check the relations. $\square$

Remark 3.3. There is some ambiguity as to whether $\mathcal{F}_V$ is meant to respect duals. *A priori*, it does not: this is because **Rib** is reflexive (double dual is the identity) but the image of $\mathcal{F}_V$ may not be; therefore we cannot expect that $\mathcal{F}_V((+)^{**}) = \mathcal{F}_V((+))^{**}$. Likewise, whereas $\widetilde{\mathrm{coev}}_{(w)} = \mathrm{coev}_{(w*)}$ in **Rib**, this will not be true in general ribbon categories. Now, note that we have shown in §2 that there *is* a canonical isomorphism between V and V^{**} for all objects V in a ribbon category; therefore, as in the case of non-strict categories, we can always pass to an equivalent category in which $V = V^{**}$. However, it is not not clear to the author how much structure is lost; this point is not mentioned in the literature. In any case, we can do everything we need to in the image of $\mathcal{F}_V$ without its domain being reflexive; we just have to be a little more careful.

Taking the previous remark into consideration, we have, after possibly restricting to an equivalent category, also called $\mathcal{C}$,

Corollary 3.1. *For any ribbon category $\mathcal{C}$ and any object $V \in \mathcal{C}$, there is a unique functor $\mathcal{F}_V$: **Rib** $\to \mathcal{C}$, respecting tensor products, duals, braidings, and ribbons, such that $\mathcal{F}_V(+) = V$.*

$\mathcal{F}_V$ is the *Reshetikhin-Turaev* functor.

Intuitively, the above discussion amounts to stating that **Rib** is the smallest ribbon category, consisting only of the minimal structure needed to make it ribbon. In fact, Corollary 3.1 allows us to represent maps of a general ribbon category as modified tangle diagrams, and there are many non-obvious relations in general ribbon categories that can be proven geometrically, with this point view. Indeed, we can see the geometric properties of ribbons as motivating many of the obstructions we went through in §2. For example, so that our duals would be tensorial, we had to define

$$\widetilde{\mathrm{ev}}_V = \mathrm{ev}_V \circ (b_{V^*}^{-1} \otimes \mathrm{id}) \circ \sigma_{V,V^*} : V \otimes V^* \to I \tag{14}$$

instead of the more naïve

$$\widetilde{\mathrm{ev}}_V = \mathrm{ev}_V \circ \sigma_{V,V^*} : V \otimes V^* \to I$$

And indeed, in the following figure, we have drawn $\widetilde{\mathrm{ev}}_+$ on the left, and (14) on the right (actually, we have written $\mathrm{ev}_+ \circ (\mathrm{id} \otimes b_-^{-1}) \circ \sigma_{+,-}$, which is equal to (14) in any ribbon category); this demonstrates geometrically what is "wrong" with omitting the ribbon element (i.e., without a ribbon element, the two sides of the equality in Figure 4 most certainly not be equal)

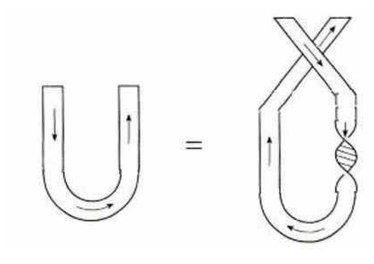

Figure 4:

Let $\mathcal{C}$ be a ribbon category with tensor identity object k. Then by Remark 3.1 and Corollary 3.1, for every object $V \in \mathcal{C}$ we get an invariant of framed links taking values in $\mathrm{End}(k)$, simply by taking the image of any link under $\mathcal{F}_V$. However, it is obvious that if the braiding of V in $\mathcal{C}$ satisfies $\sigma_{V,V}^2 = \mathrm{id}_{V,V}$, then this invariant does not distinguish any links. Therefore, we cannot apply the machinery from §2 to get link invariants unless we can find some non-trivial ribbon Hopf algebras. This we do in the next section. For now, we write down some values of $\mathcal{F}_V$ in Figure 5.

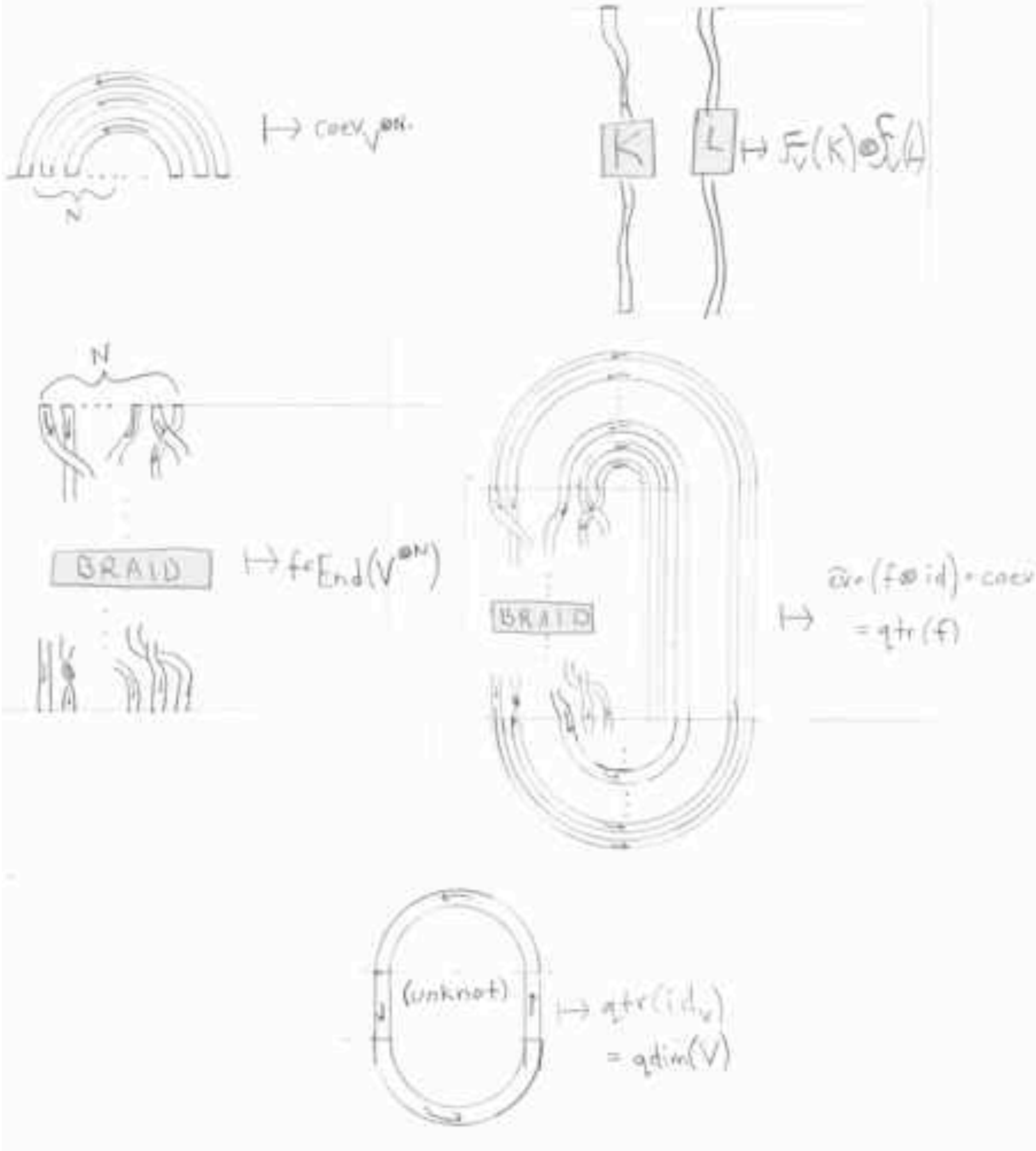
↦ coev$_V^{\otimes N}$
N
↦ $F_V(K) \otimes F_V(L)$
N
BRAID
↦ $f \in \mathrm{End}(V^{\otimes N})$
BRAID
$\widetilde{ev} \circ (f \otimes id) \circ coev$
$= qtr(f)$
(unknot)
↦ $qtr(id_V)$
$= qdim(V)$

Figure 5:

4 Quantum Groups

In this section, we construct a non-trivial ribbon Hopf algebra, generalizing the construction of $U(\mathfrak{g})$ in §2. We will be working over a ring instead of a field, but the generalization from §2 is immediate. References for this material include [10, 2, 4]. In particular, all unproved material can be found in [10].

4.1 Topological $\mathbb{C}[[h]]$-Modules

Let $\mathbb{C}[[h]]$ denote the ring of formal power series in h, with coefficients in $\mathbb{C}$. We are interested in modules over $\mathbb{C}[[h]]$. In particular, if V is a vector space over $\mathbb{C}$, then the set

$$V[[h]] = \left\{ \sum_{n=0}^{\infty} v_n h^n : v_n \in V \right\}$$

of formal power series in h with coefficients from V is a $\mathbb{C}[[h]]$-module in the obvious way, which we might call the "module quantization" of V. The action of $\mathbb{C}[[h]]$ on the quotient $V[[h]]/hV[[h]]$ naturally restricts to a $\mathbb{C}$-action, and it is easy to deduce that

$$V[[h]]/hV[[v]] \cong V$$

as vector spaces. We think of this as setting $h = 0$ to get the "classical limit:" given $a \in V[[h]]$, we write "$a \mod h$" for its image in the quotient. Likewise, given two vector spaces V, W and a $\mathbb{C}[[h]]$-linear map $\rho : V[[h]] \to W[[h]]$, we obtain a $\mathbb{C}$-linear map "$\rho \mod h$" from $V \to W$. More generally, for any $\mathbb{C}[[h]]$-module M, we define the classical limit M/hM, which *a priori* only has the structure of a $\mathbb{C}[[h]]$-module.

Returning to the module quantization $V[[h]]$, we can also take structure in the other direction, as follows. Given a general $\mathbb{C}[[h]]$-module M, any $\mathbb{C}[[h]]$-linear map $V[[h]] \to M$ is determined (by linearity) by its values on the constant power series, which form a copy of V (as an abelian group) lying in $V[[h]]$. Therefore, any $\mathbb{C}$-linear map $\rho : V \to W$ determines a unique map $\rho_h : V[[h]] \to W[[h]]$, which acts as ρ on the constant power series in $V[[h]]$. Note that $\rho_h \equiv \rho \mod h$, but of course, for any given ρ, there will be many other $\mathbb{C}[[h]]$-linear maps with this property.

Suppose that V actually has the structure (V, μ, η) of an (associative, unital) algebra over $\mathbb{C}$; we want to put a corresponding algebra structure on $V[[h]]$. By the previous discussion, we get a $\mathbb{C}[[h]]$-linear map $\mu_h : (V \otimes V)\,[[h]] \to V[[h]]$. However, in general,

$$(V \otimes V)\,[[h]] \ncong V[[h]] \otimes V[[h]], \tag{15}$$

and so μ_h does not give us a multiplication on $V[[h]]$. Now, suppose $V = \operatorname{span}\{e_i\}_{i=1}^{\infty}$; then we can think of $(V \otimes V)[[h]]$ as consisting of elements of the form

$$\sum_{n=0}^{\infty} \left(\sum_{j=0}^{m_n} e_{p_j} e_{q_j} \right) h^n,$$

where e_i, e_j are thought of as non-commuting variables. With this notation, we can embed $V[[h]] \otimes V[[h]]$ into $(V \otimes V)[[h]]$ as the sub-module spanned by power series of the form

$$\left(\sum_{n=0}^{\infty} e_{p_n} h^n \right) \cdot \left(\sum_{n=0}^{\infty} e_{q_n} h^n \right)$$

Then it is clear that every finite power series is contained in the image of $V[[h]] \otimes V[[h]]$, and therefore we can "approximate" any power series in $(V \otimes V)[[h]]$. Motivated by this example, we solve the problem presented by (15) by putting a uniform topology on $\mathbb{C}[[h]]$-modules in which $(V \otimes V)[[h]]$ is the completion of $V[[h]] \otimes V[[h]]$, and then we will define a new "topological" tensor product, which is simply the old product followed by completion in this topology. The details of this construction are not very important to what follows, so we will just give a brief overview.

Formally, we define the *h-adic* topology on a $\mathbb{C}[[h]]$-module M by declaring that an open basis of $x \in M$ consists of the sets $x + h^n M$. The completion $\varprojlim_n M/h^n M$ of M in this topology is denoted $\widehat{M}$. In the special case $M = V[[h]]$, we can define the h-adic topology as a norm topology, and the completion is just the metric-completion. Indeed, let

$$f = \sum_{n=0}^{\infty} v_n h^n, \;\; v_n \in V$$

be an arbitrary element of $V[[h]]$, and define $w(f) \in \mathbb{N} \cup \infty$ as follows: if $f \neq 0$, $w(f)$ is the smallest natural number such that $v_{w(f)} \neq 0$ but $v_k = 0$ for all $k < w(f)$; if $f = 0$, $w(f) = \infty$. Then we define a norm on $V[[h]]$ via $|f| = 2^{-w(f)}$ (it is not too hard to check that this is indeed a norm); for obvious reasons, the ultrametric defined on $V[[h]]$ by this norm is called the *h-adic metric*, and the topology induced by this metric agrees with the previous definition of the h-adic topology on $V[[h]]$. From this point of view it is obvious that an open base for zero consists of the ideals (h^n). This discussion will hopefully aid in visualizing the topology.

Define the *topological tensor product* $\widetilde{\otimes}$ between $\mathbb{C}[[h]]$-modules M and N to be the completion of the standard module tensor product:

$$M \widetilde{\otimes} N = \widehat{M \otimes N}$$

Then, by the definitions, we have

Theorem 4.1. $V[[h]] \widetilde{\otimes} W[[h]] \cong (V \otimes W)[[h]]$.

What is important to us is simply the fact that this tensor product has all the desirable properties, functoriality, associativity, etc, which the standard product enjoys, as long as we restrict ourselves to maps which our continuous in the h-adic topology. For modules of the form $V[[h]]$, it is obvious that all linear maps are continuous. Using the topological tensor product, we can generalize all of the algebra and coalgebra machinery from §2 to the case of modules over $\mathbb{C}[[h]]$. (Technically, we append this generalized machinery with the adjective "topological," e.g. "topological algebra." We will usually neglect this formality, always assuming that $\mathbb{C}[[h]]$-modules are tensored with the topological tensor product). For example, using Theorem 4.1, it's not difficult to show

Theorem 4.2. *If (A, μ, η) is a unital, associative algebra, then so is $(A[[h]], \mu_h, \eta_h)$ with the topological tensor product. Furthermore, if we set $\tilde{\mu} = \mu_h \mod h$ and $\tilde{\eta} = \eta_h \mod h$, then we have $A \cong A[[h]]/hA[[h]]$ as algebras.*

Indeed, any of the algebraic structure on A (algebra, coalgebra, Hopf algebra, braidings and ribbons), can be lifted to $A[[h]]$.

As a generalization of this "lifting" of algebraic structure, we have

Definition 4.1 (Quantization)**.** A *quantization* (sometimes called deformation) of a Hopf algebra A over $\mathbb{C}$ is a (topological) Hopf algebra A_h over $\mathbb{C}[[h]]$ such that $A_h \cong A[[h]]$ as modules, and $A_h/hA_h \cong A$ as a Hopf algebra.

By the previous discussion, we know that there is a trivial way to turn $A[[h]]$ into a quantization of A. This is called the *trivial quantization.*

Before we move on to the construction of an explicit non-trivial quantization, a few more straightforward definitions are in order. Suppose $P = \mathbb{C}\langle\{X_i\}\rangle$ is the free algebra over $\mathbb{C}$ generated by the set $\{X_i\}$; then the trivial quantization $P[[h]]$ is the called the *free topological algebra over* $\mathbb{C}[[h]]$ *generated by* $\{X_i\}$. Suppose we have a set of relations in $P[[h]]$ among these generators; let I be the two-sided ideal generated by these relations, and $\bar{I}$ its completion in the h-adic topology. Then $P[[h]]/\bar{I}$ is called the topological algebra presented by these generators and relations.

Suppose we have a $\mathbb{C}[[h]]$-algebra $V[[h]]$. Then given $v \in V$, we define

$$e^{hv} = \sum_{n=0}^{\infty} \frac{v^n h^n}{n!}$$

Likewise, we define $\sinh hv, \cosh hv$, etc. Note that e^{hv} is obviously invertible, because its inverse is e^{-hv}.

Lastly, note that the units in $\mathbb{C}[[h]]$ are precisely the power series with non-zero constant term. Therefore, given two non-zero power series $f, g \in \mathbb{C}[[h]]$ whose first non-zero coefficients have the same index, we may naturally define elements f/g and g/f by first "dividing out" and "canceling" h's. We will do this implicitly throughout.

4.2 The Quantized Universal Enveloping Algebra of $\mathfrak{sl}_2$

We are still hunting for non-trivial ribbon Hopf algebras. Unfortunately, it is obvious that the trivial quantization of $U(\mathfrak{sl}_2)$ is still cocommutative (we have just restricted to $\mathfrak{g} = \mathfrak{sl}_2$, and will continue to do so for the rest of this paper. But it is part of the richness of the subject that similar considerations work in general, at least for all semi-simple finite-dimensional Lie algebras over $\mathbb{C}$). In this section, we discuss a non-trivial quantization of $U(\mathfrak{sl}_2)$.

Definition 4.2. $U_h(\mathfrak{sl}_2)$ is the $\mathbb{C}[[h]]$-algebra with generators $\{E, F, H\}$ and relations

$$[H,E] = 2E, \quad [H,F] = -2F, \quad [E,F] = \frac{e^{hH} - e^{-hH}}{e^h - e^{-h}} = \frac{\sinh(hH)}{\sinh(h)}$$

Remark 4.1. In the literature, our h is often replaced by $h/2$.

Theorem 4.3. *$U_h(\mathfrak{sl}_2)$ has a Hopf algebra structure given on its generators by*

$$\Delta(H) = H \otimes 1 + 1 \otimes H, \quad \Delta(E) = E \otimes e^{hH} + 1 \otimes E, \quad \Delta(F) = F \otimes 1 + e^{-hH} \otimes F$$

$$\epsilon(H) = \epsilon(E) = \epsilon(F) = 0$$
$$S(H) = -H, \quad S(E) = -e^{hH}E, \quad S(F) = -e^{-hH}F$$

Sketch of Proof. Let $I \subset P[[h]]$ be the Hopf ideal generated by the relations in Definition 4.2, and let $\bar{I}$ denote its closure in the h-adic topology. To show, for example, that Δ extends to an algebra homomorphism $U_h(\mathfrak{sl}_2) \to U_h(\mathfrak{sl}_2) \otimes U_h(\mathfrak{sl}_2)$, we need to show that $\Delta(I) \subset P[[h] \otimes I + I \otimes P[[h]]$ on the (algebra) basis of $P[[h]]$; it would then follow that $\Delta(\bar{I}) \subset P[[h] \otimes \bar{I} + \bar{I} \otimes P[[h]]$ by the $\mathbb{C}[[h]]$-linearity of Δ. The former statement follows from a straightforward computation. Checking similar statements for ϵ and S is not difficult; furthermore, it is easy to check that the Hopf algebra relations are satisfied. □

Theorem 4.4. *$U_h(\mathfrak{sl}_2)$ is a quantization of $U(\mathfrak{sl}_2)$. It is known as the Quantized Universal Enveloping Algebra of $\mathfrak{sl}_2$.*

It is easy to check from the definitions that $U_h(\mathfrak{sl}_2)/hU_h(\mathfrak{sl}_2) \cong U(\mathfrak{sl}_2)$ as a Hopf algebra. It remains to show that $U_h(\mathfrak{sl}_2) \cong U(\mathfrak{sl}_2)[[h]]$ as $\mathbb{C}[[h]]$-modules. In fact, we show that the entire algebra structure is preserved:

Theorem 4.5. *$U_h(\mathfrak{sl}_2) \cong U(\mathfrak{sl}_2)[[h]]$ as algebras over $\mathbb{C}[[h]]$.*

On the other hand, the coalgebra structure is deformed; in particular, it is obvious that $U_h(\mathfrak{sl}_2)$ is not cocommutative.

During the proof, we will distinguish elements of $U(\mathfrak{sl}_2)$ from their counterparts in $U_h(\mathfrak{sl}_2)$ by denoting the classical elements with bars. Following [2], we will prove Theorem 4.5 by constructing an explicit algebra isomorphism $\varphi : U_h(\mathfrak{sl}_2) \cong U(\mathfrak{sl}_2)[[h]]$. To do this, we make use of a quantum analogue of the Casimir element

$$\bar{\Omega} = \frac{1}{4}(\bar{H}+1)^2 + \bar{F}\bar{E} = \frac{1}{4}(\bar{H}-1)^2 + \bar{E}\bar{F},$$

which generates the center of $U(\mathfrak{sl}_2)$. This element will also be important for the ribbon structure on $U_h(\mathfrak{sl}_2)$.

Definition 4.3. The *quantum Casimir element* of $U_h(\mathfrak{sl}_2)$ is

$$\Omega = \left(\frac{\sinh \frac{1}{2}h(H+1)}{\sinh h}\right)^2 + FE = \left(\frac{\sinh \frac{1}{2}h(H-1)}{\sinh h}\right)^2 + EF$$

It is easy to check that the two expressions above are indeed equal, and also that $\bar{\Omega} \equiv \Omega \mod h$.

Proof of Theorem 4.5. We define $\varphi : U_h(\mathfrak{sl}_2) \to U(\mathfrak{sl}_2)[[h]]$ on the generators of $U_h(\mathfrak{sl}_2)$ as follows:

$$\varphi(H) = \bar{H},\ \varphi(F) = \bar{F}$$
$$\varphi(E) = 2\left(\frac{\cosh h(\bar{H}-1) - \cosh 2h\sqrt{\bar{\Omega}}}{((\bar{H}-1)^2 - 4\bar{\Omega})\sinh^2 h}\right)\bar{E}$$

To see that the above quotient is well-defined, note that, for indeterminates u, v, the expression $\frac{\cosh u - \cosh v}{u^2 - v^2}$ can be written as a formal power series $f(u^2, v^2)$. We can also see that $\varphi \equiv \mathrm{id} \mod h$.

To show that φ extends from the basis to a homomorphism of algebras, the only non-trivial identity to check is

$$\varphi(E)\bar{F} - \bar{F}\varphi(E) = \frac{\sinh h\bar{H}}{\sinh h} \tag{16}$$

(we have already substituted $\varphi(H) = \bar{H}$ and $\varphi(F) = \bar{F}$).

Substitute the definition of $\bar{\Omega}$ into the quotient to get

$$\varphi(E)\bar{F} = -\left(\frac{\cosh h(\bar{H}-1) - \cosh 2h\sqrt{\bar{\Omega}}}{2\sinh^2 h}\right)$$

Furthermore, because $\bar{F}(\bar{H}-1) = (\bar{H}+1)\bar{F}$, and because $\bar{\Omega}$ is in the center of $U(\mathfrak{sl}_2)$, we have

$$\bar{F} f\left((h^2(\bar{H}-1)^2, 4h^2\bar{\Omega}\right) = f\left(h^2(\bar{H}+1)^2, 4h^2\bar{\Omega}\right)\bar{F}$$

(f is the formal power series just defined), so that

$$\bar{F}\varphi(E) = -\left(\frac{\cosh h(\bar{H}+1) - \cosh 2h\sqrt{\bar{\Omega}}}{2\sinh^2 h}\right)$$

Subtracting gives (16).

To prove that φ is an isomorphism, we construct its inverse. First, reusing some of the previous computation, it is easy to derive

$$\varphi(\Omega) = \frac{\sinh^2 h\sqrt{\bar{\Omega}}}{\sinh^2 h}$$

Letting

$$g(h,u) = \frac{\sinh^2 h\sqrt{u}}{\sinh^2 h} \in \mathbb{C}[[u,h]],$$

it is not hard to show

Lemma 4.1. *There exists $\tilde{g} \in \mathbb{C}[[h,u]]$ such that*

$$g(h, \tilde{g}(h,u)) = \tilde{g}(h, g(h,u)) = u$$

Now define $\psi(\bar{H}) = H,\ \psi(\bar{F}) = F$, and

$$\psi(\bar{E}) = \frac{1}{2}\left(\frac{((H-1)^2 - 4\tilde{g}(h,\Omega))\sinh^2 h}{\cosh h(H-1) - \cosh 2h\sqrt{\tilde{g}(h,\Omega)}}\right)E$$

It is not hard to compute $\psi(\varphi(E)) = E,\ \varphi(\psi(\bar{E})) = \bar{E}$, so we will be finished once we show that ψ extends to an algebra homomorphism. Again, we just compute

$$\begin{aligned}\psi(\bar{E})F \quad &- \quad F\psi(\bar{E}) \\ &= \frac{1}{2}\left(\frac{((H-1)^2 - 4\tilde{g}(h,\Omega))\sinh^2 h}{\cosh h(H-1) - \cosh 2h\sqrt{\tilde{g}(h,\Omega)}}\right)EF \\ &- \frac{1}{2}\left(\frac{((H+1)^2 - 4\tilde{g}(h,\Omega))\sinh^2 h}{\cosh h(H+1) - \cosh 2h\sqrt{\tilde{g}(h,\Omega)}}\right)FE\end{aligned}$$

where we have used $[H,F]=-2E$ to bring the F to the right side in the second term above. Using the definition of Ω, we then rewrite the above as

$$\begin{aligned} & \frac{1}{2}\left(\frac{((H-1)^2-4\tilde{g}(h,\Omega))\sinh^2 h}{\cosh h(H-1)-\cosh 2h\sqrt{\tilde{g}(h,\Omega)}}\right)\left(\Omega-\frac{\sinh^2\frac{1}{2}h(H-1)}{\sinh^2 h}\right) \\ - & \frac{1}{2}\left(\frac{((H+1)^2-4\tilde{g}(h,\Omega))\sinh^2 h}{\cosh h(H+1)-\cosh 2h\sqrt{\tilde{g}(h,\Omega)}}\right)\left(\Omega-\frac{\sinh^2\frac{1}{2}h(H+1)}{\sinh^2 h}\right) \end{aligned}$$

Now, we simply transform the above denominator as follows:

$$\begin{aligned} \cosh h(H-1) & - \cosh 2h\sqrt{\tilde{g}(h,\Omega)} \\ & = 2\left(\sinh^2\frac{1}{2}h(H-1)-\sinh^2 h\sqrt{\tilde{g}(h,\Omega)}\right) \\ & = 2\left(\sinh^2\frac{1}{2}h(H-1)-\Omega\sinh^2 h\right) \end{aligned}$$

Therefore

$$\begin{aligned} \psi(\bar{E})F & - F\psi(\bar{E}) \\ & = \frac{1}{4}((H+1)^2-4\tilde{g}(h,\Omega))-\frac{1}{4}((H-1)^2-4\tilde{g}(h,\Omega)) \\ & = H \end{aligned}$$

□

Remark 4.2. Using φ, one can show that Ω is a "topological" generator for the center of $U_h(\mathfrak{sl}_2)$ (i.e., it generates a dense subset of the center).

To derive knot invariants from $U_h(\mathfrak{sl}_2)$, we will need to know something about its representations (and that it is a ribbon algebra, of course, which we'll see shortly). In fact, Theorem 4.5 tells us almost everything: we can derive the representation theory of $U_h(\mathfrak{sl}_2)$ directly from that of $\mathfrak{sl}_2$, as we see now.

First, define $\mathbf{Rep}_A^{fin}$, in the case of a topological algebra A, to be the category of representations which are free and finite rank as $\mathbb{C}[[h]]$-modules. Now, suppose A is a quantization, i.e., there exists an algebra K over $\mathbb{C}$ such that $A\cong K[[h]]$ as algebras. Then a representation $V\in\mathbf{Rep}_A^{fin}$ produces a representation $V/hV\in\mathbf{Rep}_K^{fin}$. Likewise, given an object $V\in\mathbf{Rep}_K^{fin}$, we know how to produce an object $V[[h]]\in\mathbf{Rep}_A^{fin}$. In the case that $A=U_h(\mathfrak{sl}_2)$, it is known that these operations are mutually inverse, because the finite-dimensional representations of $\mathfrak{sl}_2$ admit no non-trivial deformations (see [13, 10, 2]). Finally, irreducible representations of K correspond to "indecomposable" representations of A, where an indecomposable representation $V[[h]]$ of A, for $V\in\mathbf{Rep}_K^{fin}$, is one which cannot be written as the sum of representations of the form $W[[h]], U[[h]]$ for $W,U\in\mathbf{Rep}_K^{fin}$ (in the stricter sense, any representation $V\in\mathbf{Rep}_A^{fin}$ has a proper subrepresentation hV). Returning to $U_h(\mathfrak{sl}_2)$, these indecomposable representations actually have quite a beautiful form, which the reader can compute directly from the well-known representation theory of $U(\mathfrak{sl}_2)$, now that we've produced an isomorphism $\varphi : U_h(\mathfrak{sl}_2) \to U(\mathfrak{sl}_2)[[h]]$; we will just write them down without deriving them. First, we need to define the *quantum integer*

$$[n] = \frac{q^n - q^{-n}}{q - q^{-1}}$$

Then we can also define the quantum factorial

$$[n]! = [n] \cdot [n-1] \cdots [2] \cdot [1], \quad \frac{[n]!}{[n-m]!} = [n] \cdot [n-1] \cdots [n-m+1]$$

which will come up later. For the rest of this chapter, set $q = e^h$. Then it is easy to check that $[n]$ is a well-defined element of $\mathbb{C}[[h]]$, and that $[n] \equiv n \mod h$; likewise for the quantum factorial (we may sometimes write $[n]_q$ instead of $[n]$; it is the same thing). Then we have

Theorem 4.6. *Just as for $U(\mathfrak{sl}_2)$, the finite dimensional, indecomposable representation for $U_h(\mathfrak{sl}_2)$ are indexed by the positive integers. Explicitly, on generators, the nth representation V_n, into the free module of rank $n+1$ over $\mathbb{C}[[h]]$ with basis $\{v_0, \ldots, v_n\}$, takes the values*

$$H.v_r = (n-2r)v_r \tag{17}$$
$$E.v_r = [n-r+1]v_{r-1} \tag{18}$$
$$F.v_r = [r+1]v_{r+1} \tag{19}$$

Remark 4.3. If we use new generators $X = Ee^{-hH/2}$ and $Y = e^{hH/2}F$, we can rewrite the above representations in matrix form as

$$\rho_n(X) = \begin{pmatrix} 0 & [n]_q & 0 & \ldots & 0 \\ 0 & 0 & [n-1]_q & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & \ldots & 0 & 0 \end{pmatrix} \rho_n(Y) = \begin{pmatrix} 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & [2]_q & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & \ldots & [n]_q & 0 \end{pmatrix}$$

$$\rho_n(H) = \begin{pmatrix} n & 0 & \ldots & 0 & 0 \\ 0 & n-2 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & -n+2 & 0 \\ 0 & 0 & \ldots & 0 & -n \end{pmatrix}$$

which is the form that appears most often in the literature. Indeed, for the rest of this paper, we will use the notation $\rho_n : U_h(\mathfrak{sl}_2) \to V_n$ to denote the indecomposable, $n+1$-dimensional representation of $U_h(\mathfrak{sl}_2)$, and we will continue to fix a preferred basis $\{v_0, \cdots, v_n\}$ of V_n.

It is quite straightforward to check that the above maps really do define representations of $U_h(\mathfrak{sl}_2)$, and also to compute that V_n/hV_n is the $n+1$-dimensional irreducible representation of $U(\mathfrak{sl}_2)$.

We now produce a universal R-matrix and ribbon element for $U_h(\mathfrak{sl}_2)$, turning it into a ribbon Hopf algebra. Historically, the existence of a universal R-matrix was one of the most exciting features of $U_h(\mathfrak{s}l_2)$, for topologists, and for those interested in statistical mechanics and quantum integrable systems. In fact, Drinfel'd proved that there is a universal R-matrix not only for $U_h(\mathfrak{s}l_2)$, but for many other Quantized Universal Enveloping Algebras as well. To do so, he made use of a novel construction called the *quantum double*: he showed, in quite a general setting, that if two Hopf algebras A, B are dual in an appropriate sense, say for example, that we have a pairing $e_i \mapsto e^i$ between their bases with certain properties, then one can create a new Hopf algebra, essentially of the form $\mathcal{D} = A \otimes B^{op}$. He then showed that the unique element $\sum e_i \otimes e^i \in \mathcal{D}$ was a universal R-matrix for $\mathcal{D}$.

In fact, for $U_h(\mathfrak{s}l_2)$, the roles of A and B are essentially played by $U_h(\mathfrak{b}^+)$ and $U_h(\mathfrak{b}^-)$, the quantizations of the positive and negative Borel subalgebras of $\mathfrak{s}l_2$. However, the situation turns out to be slightly tricky; the full story is given in the references mentioned at the beginning of the section. Nevertheless, using the general form of the R-matrix suggested by Drinfeld's formalism, it is not so hard to compute the coefficients which ensure that $U_h(\mathfrak{s}l_2)$ is quasitriangular; this is done in [12]. Here, we simply produce the fabled universal R-matrix:

Theorem 4.7. *$U_h(\mathfrak{s}l_2)$ is topologically quasi-triangular, with universal R-matrix*

$$\mathcal{R} = \sum_{n=0}^{\infty} \frac{q^{\frac{1}{2}n(n-1)}(q-q^{-1})^n}{[n]!} e^{\frac{1}{2}h(H\otimes H)} (E)^n \otimes (F^n) \tag{20}$$

Recall that, according to §2, $\tau \circ \rho_n(\mathcal{R})$ defines a $U_h(\mathfrak{s}l_2)$-linear map $V_n \otimes V_n \to V_n \otimes V_n$. Denote this map by $\check{R}_n$, and let $(\check{R}_n)^{ij}_{kl}$ be the $v_k \otimes v_l$-coordinate of $\check{R}_n(v_i \otimes v_j)$. From Theorem 4.7, it is not difficult to compute (see [12]):

Theorem 4.8.

$$\begin{aligned}(\check{R}_n)^{ij}_{kl} = \sum_{n=0}^{\min(n-1-i,j)} & \delta_{l,i+n}\delta_{k,j-n} \frac{(q-q^{-1})^n}{[n]!} \frac{[i+n]!}{[i]!} \frac{[n-1+n-j]!}{[n-1-j]!} \\ & \times q^{2(i-(n-1)/2)(j-(n-1))/2-n(i-j)-n(n+1)/2}\end{aligned}$$

Now, let $u = m\left((S \otimes \mathrm{id})(\tau(\mathcal{R}))\right) = S(r_{(2)})r_{(1)}$ (with $\mathcal{R} = \sum r_{(1)} \otimes r_{(2)}$), as in (9), so that $S^2(a) = uau^{-1}$ (see (12)) for all $a \in U_h(\mathfrak{s}l_2)$. It is quite an easy computation on the generators of $U_h(\mathfrak{s}l_2)$ to check directly that $S^2(a) = e^{hH}ae^{-hH}$ as well. Therefore $\nu = e^{-hH}u = ue^{-hH}$ is in the center of $U_h(\mathfrak{s}l_2)$; because e^{hH} satisfies $\Delta(e^{hH}) = e^{hH} \otimes e^{hH}$, we conclude that ν is a ribbon element of $U_h(\mathfrak{s}l_2)$. Furthermore, Drinfeld has shown the following nice realization of ν (see [3]):

Theorem 4.9. $\nu = e^{\frac{h\Omega}{2}}$.

Furthermore, we have $\kappa = \nu^{-1}u = e^{hH}$, so that

$$\rho_n(\kappa) = \begin{pmatrix} q^n & 0 & \dots & 0 & 0 \\ 0 & q^{n-2} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & q^{-n+2} & 0 \\ 0 & 0 & \dots & 0 & q^{-n} \end{pmatrix} \tag{21}$$

Likewise, by computing $\rho_n(\Omega)$, and using Theorem 4.9, we have

$$\rho_n(\nu^{-1}) = q^{-\frac{(n+1)^2-1}{2}} \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} \tag{22}$$

Indeed, because ν is central and V_n indecomposable, it can be deduced *a priori* that $\rho_n(\nu)$ gives a scalar operator. (In fact, $\rho_n(\Omega)$ gives the same operator as the classical Casimir element).

We now proceed to discuss the topological invariants derived from $U_h(\mathfrak{sl}_2)$.

5 Quantum Knot Invariants

5.1 The Jones Polynomial

As mentioned, because V_n is indecomposable and the ribbon element ν^{-1} is central, its image in $\mathrm{End}(V_n)$ is a scalar operator; let α_n denote this scalar. We first show that this allows us to get an unframed invariant from $\mathcal{F}_{V_n}$.

Let T be an oriented tangle, not ribbon, i.e. an unframed tangle. Given any diagram D of T, there is an obvious framing, called the blackboard framing, where the ribbon is simply taken to be flush against the plane of projection, which turns D into a ribbon $b(D)$. We also let $w(D)$ denote the *writhe* of D, which is defined to be the sum of the signs of the crossings of D, see Figures 6 or 7.

Theorem 5.1. *$F_n(T) := \alpha_n^{-w(D)}\mathcal{F}_{V_n}(b(D))$ is an invariant of oriented tangles.*

Proof. We know that $\mathcal{F}_{V_n}$ is invariant under Reidemeister moves 2 and 3, so we need to check move 1. Applying move 1 to D and then taking $b(D)$ is the equivalent of adding or removing a twist, which, by the construction in §3, is the equivalent of composing $\mathcal{F}_{V_n}(b(D))$ with $\rho_n(\nu^{-1})$ or its inverse, that is, multiplying by $\alpha_n^{\pm 1}$. It is easy to check that, under such a move, the writhe of D changes by ± 1 as well. □

Definition 5.1. $F_n(R)$ is called the unframed *colored Jones polynomial* with color n.

Therefore, for an oriented (n,m)-tangle R, $F_n(R)$ is a linear operator from $\otimes_{i=1}^n V^{\epsilon_i} \to \otimes_{j=1}^m V^{\epsilon_j}$, where $\epsilon = \pm 1$ and $V^1 = V, V^{-1} = V^*$. When $n = m = 0$, so that R is an oriented link, we have $F_n(R) \in \mathrm{End}(\mathbb{C}[[h]])$, so by evaluating at 1, we get an element of $\mathbb{C}[[h]]$ which we will still refer to as $F_n(R)$. As we have seen, when we write the structure maps of $\mathbf{Rep}^{fin}_{U_h(\mathfrak{sl}_2)}$ as matrices in our fixed basis, their entries live in $\mathbb{C}(q)$, so we conclude that $F_n(R) \in \mathbb{C}(q)$ as well, when R is a link.

Given that most people are introduced to the Jones polynomial, and its immense importance, via a "skein theoretic" definition in terms of knot diagrams, the next result should seem rather stupendous:

Theorem 5.2. *When L is a link, $F_1(L)$ gives the Jones polynomial.*

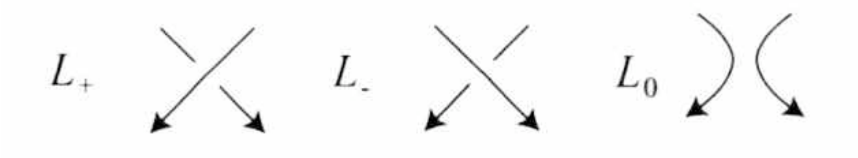

Figure 6:

Proof. It is well known (the argument is completely combinatorial) that the Jones polynomial is characterized by its value on the unknot, usually taken to be $q+q^{-1}$ or 1, and by the skein relation

$$q^2 J(L_+) + q^{-2} J(L_-) = (q - q^{-1}) J(L_0) \tag{23}$$

Using Theorem 4.6 and 4.7 or 4.8, we compute

$$\check{R}_1 = \tau \circ \rho_1(\mathcal{R}) = \begin{pmatrix} q^{1/2} & 0 & 0 & 0 \\ 0 & q^{1/2} - q^{-3/2} & q^{-1/2} & 0 \\ 0 & q^{-1/2} & 0 & 0 \\ 0 & 0 & 0 & q^{1/2} \end{pmatrix}$$

Comparing with §3, we see that this is $\mathcal{F}_{V_1}(L_+)$; we also see that L_- is sent to its inverse, which is

$$\check{R}_1^{-1} = \begin{pmatrix} q^{-1/2} & 0 & 0 & 0 \\ 0 & 0 & q^{1/2} & 0 \\ 0 & q^{1/2} & q^{-1/2} - q^{3/2} & 0 \\ 0 & 0 & 0 & q^{-1/2} \end{pmatrix}$$

Subtracting, we get

$$q^{1/2} L_+ - q^{-1/2} L_- = (q - q^{-1}) \mathrm{Id}$$

Now, by (22), $\alpha_1 = q^{-3/2}$; therefore, because $w(L_+) = w(L_0) + 1$, $w(L_-) = w(L_0) - 1$, we derive (23). Finally, we compute (by O we mean the unknot)

$$F_1(O) = \mathrm{qdim}(V_1) = \mathrm{tr}(\rho_1(\kappa)) = [2] = q + q^{-1} \tag{24}$$

by (21) and the results of §3, agreeing with a common normalization of the Jones polynomial. □

What's more, it is easy to generalize the computation in (24), giving us

$$F_n(O) = \mathrm{qdim}(V_n) = \mathrm{tr}(\rho_n(\kappa)) = [n+1],$$

nicely generalizing the classical dimension.

Remark 5.1. It can be shown, essentially because the representations of $U_h(\mathfrak{sl}_2)$ are self-dual, that the colored Jones polynomial is essentially independent of orientation.

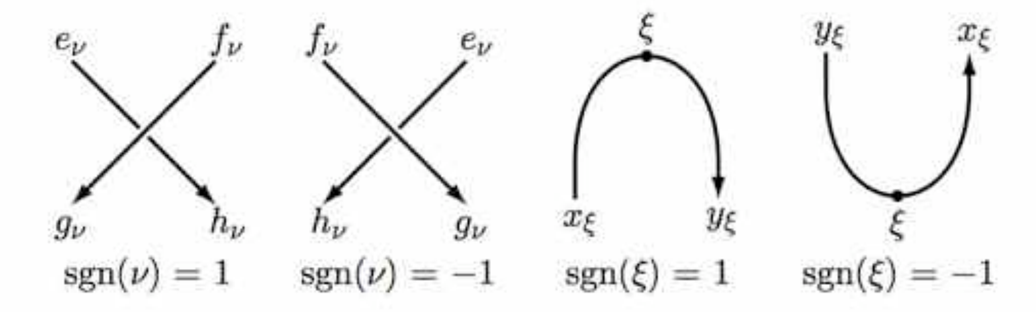

Figure 7: Signs for vertices

5.2 State-Sum Models

Now, suppose L is an oriented link, and let $D \subset \mathbb{R}^2$ be a diagram for L. Thinking of $(1,0)$ as a directed vector on the x axis in $\mathbb{R}^2$, let $\mathcal{X}$ denote the set of points of D whose tangent vectors are parallel to $(1,0)$, and $\dot{\mathcal{X}} \subset \mathcal{X}$ those points of $\mathcal{X}$ whose tangent vectors also point in the same direction as $(1,0)$. Let $\mathcal{N}$ denote the crossings of D. Consider D as a graph with vertices $\mathcal{X} \cup \mathcal{N}$, and let $\mathcal{E}$ denote the set of edges of D. Define signs of the vertices as in Figure 7.

Let Ψ be the set of functions $\sigma : \mathcal{E} \to \{1, \ldots, n+1\}$, called *states*. According to Figure 7, if $\sigma \in \Psi$ and $\nu \in \mathcal{N}$, define

$$\langle D|\sigma\rangle_\nu = \begin{cases} (\check{R}_n)^{\sigma(e_\nu)\sigma(f_\nu)}_{\sigma(g_\nu)\sigma(h_\nu)} & \text{if } \mathrm{sgn}(\nu) = 1 \\ (\check{R}_n^{-1})^{\sigma(f_\nu)\sigma(e_\nu)}_{\sigma(h_\nu)\sigma(g_\nu)} & \text{if } \mathrm{sgn}(\nu) = -1 \end{cases} \tag{25}$$

Lastly, let μ^i_j be i,j entry of the $n+1 \times n+1$ matrix $\rho_n(\kappa)$, in our preferred basis. Then we have

Theorem 5.3.

$$F_n(L) = \alpha_n^{-w(L)} \sum_{\sigma\in\Psi} \left(\prod_{\nu\in\mathcal{N}} \langle D|\sigma\rangle_\nu \prod_{\xi\in\dot{\mathcal{X}}} (\mu^{\sigma(x_\xi)}_{\sigma(y_\xi)})^{-\epsilon(\xi)} \prod_{\xi\in\mathcal{X}-\dot{\mathcal{X}}} \delta_{\sigma(x_\xi),\sigma(y_\xi)} \right) \tag{26}$$

Proof. See figure 8 to recall how $\mathcal{F}_{V_n}$ acts on the elementary tangles (with the blackboard framing).

If, after dividing D into elementary tangles, only those in Figure 8 appear, then the theorem is obvious; it follows from a straight-forward linear algebra computation. One would hope that, in the more general case, the situation would be similar; however, things are surprisingly subtle. In [24, p. 540], Turaev, after arguing in a special case as above, simply shows that the general expression in (26) is invariant under Reidemeister moves. We refer the reader to his paper for the proof. $\square$

Remark 5.2. Given that $\rho_n(\kappa)$ is a diagonal matrix, (26) can be written in the more symmetric form

$$F_n(L) = \alpha_n^{-w(L)} \sum_{\sigma\in\Psi} \left(\prod_{\nu\in\mathcal{N}} \langle D|\sigma\rangle_\nu \prod_{\xi\in\dot{\mathcal{X}}} \mu^{-\epsilon(\xi)}_{\sigma(y_\xi)} \delta_{\sigma(x_\xi),\sigma(y_\xi)} \prod_{\xi\in\mathcal{X}-\dot{\mathcal{X}}} \delta_{\sigma(x_\xi),\sigma(y_\xi)} \right) \tag{27}$$

where μ_i is the ith entry along the diagonal of $\rho_n(\kappa)$.

Figure 8: Elementary tangles

To derive some geometric meaning from $F_n(L)$, it turns out we will need to specialize the variable q to particular values of $\mathbb{C}$. Note that this cannot be done *a priori*, while still in $U_h(\mathfrak{sl}_2)$; many power series in h would not converge. However, when L is a link, as we have seen, $F_n(L)$ is at least a rational function in q, and we can specialize.

For the rest of this paper, we set $q = e^{2\pi i/(n+1)}$. Immediately, this specialization appears to be a little worrisome: for example, we have $F_n(O) = [n+1] = 0$. Indeed, $F_n(L) = 0$ for all links at this value of q. To see why this might be so, note that to any link L and choice of point on L, we can associate a (1,1)-tangle simply by breaking L at that point. Denote such a tangle by L' (conversely, given a (1,1)-tangle L', we can close it to get a link, uniquely determined by L', called L). The invariant $F_n(L')$ is an operator from $V_n \to V_n$ or $V_n^* \to V_n^*$; because V_n is indecomposable, it's not hard to show that this operator is a scalar operator, see [12, p. 499]. Let $F_n(L')$ denote this scalar. By the calculations in Figure 5, we have

$$\begin{aligned} F_n(L) &= \operatorname{qtr}(F_n(L')) \\ &= \operatorname{tr}\left(\rho_N(\kappa \circ \nu^{-1}) \circ \left(F_n(L') \cdot \mathrm{Id}_{V_N}\right)\right) \\ &= F_n(L')\operatorname{tr}(\rho(\kappa \circ \nu^{-1})) \\ &= F_n(L') \cdot \operatorname{qdim}(V_N) \end{aligned}$$

Thus we define

$$F'_n(L) = F_n(L') = \frac{F_n(L)}{[n+1]}$$

(of course, formally, the last expression is not well-defined when $q = e^{2\pi i/(n+1)}$; we simply use the the middle expression (or we can think of ourselves as dividing before specializing)). Note that this is the invariant to be used in the Volume Conjecture.

Suppose we have a diagram D of an oriented (1,1)-tangle L'; D corresponds to a link diagram with two distinguished edges, adjacent to the broken point. Consider D as a graph as before, and let $\dot{\Psi}_A$ be the set of states which take some fixed $A \in \{1, \ldots, n+1\}$ on these two distinguished edges. Then a simple matrix computation generalizes Theorem 5.3 to the case of (1,1)-tangles:

Theorem 5.4. $F'_n(L) = \alpha_n^{-w(L)} \sum_{\sigma \in \dot{\Psi}_A} \left(\prod_{\nu \in \mathcal{N}} \langle D|\sigma \rangle_\nu \prod_{\xi \in \dot{\mathcal{X}}} (\mu_{\sigma(y_\psi)}^{\sigma(x_\psi)})^{-\epsilon(\xi)} \prod_{\xi \in \mathcal{X} - \dot{\mathcal{X}}} \delta_{\sigma(x_\xi), \sigma(y_\xi)} \right)$

(Precisely, the choice of A corresponds to which diagonal entry of the scalar operator $F'_N(L')$ we are choosing; of course, they are all the same. Without loss of generality, one usually takes $A = 0$, and we write $\dot{\Psi} = \dot{\Psi}_0$).

5.3 Kashaev's Invariant

In a series of papers [8, 9, 7], Kashaev derived a link invariant from quantum groups in a fashion similar to the one we have just gone through, though he was motivated by a quantum analogue of a complex function called the *dilogarithm*, which we will define in §6. He also introduce a state-sum model for his invariant. Later, in [17], Murakami and Murakami showed that, at $q = e^{2\pi i/(n+1)}$, one could apply some linear algebra manipulations to deduce that Kashaev's invariant agrees with $F'_n(L)$. We will need this form of the invariant in §7, so we write it down now, though we will recall it then as well.

Note that, to attempt to agree with the literature, we will have to make a somewhat radical change in notation. First, we now let $N = n + 1$; thus N is the dimension of the representation of $U_h(\mathfrak{sl}_2)$ which is lurking in the background. For $m \in \mathbb{Z}$, let $[m] \in \{1, \dots N\}$ be the residue of m. Let $(w)_{[m]} = (1 - w)(1 - w^2) \dots (1 - w^{[m]})$, and define the symbol

$$\theta^{ij}_{kl} = \begin{cases} 1 & \text{if } [i-j] + [j-l] + [l-k-1] + [k-i] = N-1, \\ 0 & \text{otherwise.} \end{cases}$$

and, for convenience, let $\bar{q} = q^{-1}$. Define the $N^2 \times N^2$ matrix S such that

$$(S)^{ij}_{kl} = N q^{-1/2-(k-j)(i-l+1)} \frac{\theta^{ij}_{kl}}{(\bar{q})_{i-j}(q)_{j-l}(\bar{q})_{l-k-1}(\bar{q})_{k-i}}$$

Remark 5.3. Note that $\theta = 1$ if and only if q^i, q^j, q^l, q^k go around the unit circle clockwise, and $l \neq k$.

Theorem 5.5 (Kashaev's version of the colored Jones polynomial). *Replacing $\check{R}$ by S in the notation from Theorem 5.3, we have*

$$F_{N-1}(L') = \prod_{\xi \in \mathcal{X}} -q^{sgn(\xi)/2} \cdot \sum_{\sigma \in \dot{\Psi}} \left(\prod_{\nu \in \mathcal{N}} \langle D|\sigma \rangle_\nu \prod_{\xi \in \mathcal{X}} \delta_{\sigma(x_\xi)+1, \sigma(y_\xi)} \right)$$

Proof. See [17]. The proof consists of some complicated matrix manipulations (note that we use the notation from [29]). □

6 Hyperbolic Geometry

We have now given a rapid overview of one route to the so-called quantum knot invariants. In this section, we make a sudden change in direction, and describe the geometry of hyperbolic 3-space. In the next and final section, we will show how the two are related. Some good references for this material are [1, 20, 16, 23].

6.1 Preliminaries

We use the symbol $\mathbb{H}^3$ to denote hyperbolic 3-space as an abstract Riemannian 3-manifold, up to isometric diffeomorphism (isometry, for short). For our purposes, it will be useful to have two different concretely-realized models of $\mathbb{H}^3$. We can use either one to define $\mathbb{H}^3$, and then derive the other via an isometry; we choose to begin with the Poincaré ball model.

Definition 6.1. The Poincaré ball B^3 is the open unit ball $B^3 \subset \mathbb{R}^3$ with Riemannian metric

$$ds_x^2 = \frac{4|dx|^2}{(1-|x|^2)^2}$$

at each point $x \in B^3$.

Definition 6.2. $\mathbb{H}^3$ is the isometry class of B^3.

Of course, B^3 is *not* isometric to the unit ball with the metric induced from $\mathbb{R}^3$ (though the two are diffeomorphic via the identity map). However, because ds_x^2 is just a positive multiple of the induced metric, B^3 is "conformally correct," that is, the standard Euclidean measure of angle agrees with the hyperbolic angle. On the other hand, distances are warped in the ball model—in particular, there are hyperbolic geodesics which are not Euclidean lines. To see this, we first describe the (orientation-preserving) isometries of B^3.

Consider the 3-sphere S^3 as $\mathbb{R}^3$ union a point at infinity, and define a *Euclidean similarity* S of S^3 to be a map of the form

$$S(p) = \lambda A(p) + b, S(\infty) = \infty$$

with $\lambda > 0$, $A \in O(3)$, $p, b \in \mathbb{R}^3$. Define the *fundamental reflection* $J : S^3 \to S^3$ to be the map

$$J(p) = \frac{p}{|p|^2}, J(0) = \infty, J(\infty) = 0$$

with $p \in \mathbb{R}^3$. Then we have

Definition 6.3. A *Möbius transformation* of S^3 is an orientation-preserving diffeomorphism $S^3 \to S^3$ obtained as the composition of a finite number of Euclidean similarities and fundamental reflections.

Möbius transformations of S^3 have much in common with their better-known relatives acting on S^2: they are conformal at points not sent to ∞, preserving Euclidean angles, and they send circles and lines to circles and lines.

For any subset $E \subset S^3$, we denote by Möb(E) the group of Möbius transformations sending E to itself. Then the following important theorem is standard (we implicitly restrict Möb(B^3) from S^3 to B^3):

Theorem 6.1. *Möb*$(B^3) =$ *Isom*$^+(B^3)$

See[16] for details.

Note that each element of Möb(B^3) restricts to a Möbius transformation of S^2, and that this restriction determines the original element of Möb(B^3) uniquely. Furthermore, every Möbius transformation of S^2 can be obtained in this way, and therefore we obtain an isomorphism Isom$^+(B^3) =$ Möb$(B^3) \cong PSL_2(\mathbb{C})$. The element of Isom$^+(B^3)$ associated to an element $\gamma \in PSL_2(\mathbb{C})$ by this isomorphism is sometimes called the *Poincaré extension* of γ.

Remark 6.1. So far, $S^2 = \partial B^3$ is only defined in terms of the special properties of the ball model. However, we will soon give an intrinsic definition of $\partial\mathbb{H}^3$ in terms of the geometry of $\mathbb{H}^3$.

Returning now to the geodesics of B^3, one first computes directly that the distance-minimizing path from the origin to any point $q \in B^3$ is given by a Euclidean line. In other words, for any path C joining 0 and q, one verifies the inequality

$$\int_C \frac{|dp|}{1-|p|^2} \geq \int_0^{|q|} \frac{dx}{1-x^2},$$

and checks that equality holds when C is the Euclidean line joining 0 to q. From this it follows that the Euclidean diameters of B^3 are geodesics. One then checks, from the circle-mapping and conformal properties of $\text{Möb}(B^3)$, that the orbits under $\text{Isom}^+(B^3)$ of these diameters consist precisely of the intersections $B^3 \cap S$, where S is a circle in $\mathbb{R}^3$ which is perpendicular to ∂B^3. Thus we have determined the geodesics of B^3. Furthermore, we define a *hyperbolic arc* to be a closed connected subset of a hyperbolic geodesic (they play the role of line segments), and we define a *hyperbolic plane* to be a closed subspace of $H \subset \mathbb{H}^3$ such that the geodesic between any two points of H is contained in H. One can then check that the hyperbolic planes of B^3 are intersections $B \cap B^3$, where $B \subset \mathbb{R}^3$ are spheres orthogonal to ∂B^3, or else planes through the origin (i.e. spheres through infinity).

Integrating along diameters of B^3, one sees that all geodesics are infinitely long, and therefore that B^3 is complete. In particular, $S^2 = \partial B^3$ is "infinitely far away" from any point in B^3. Indeed, this so-called *sphere at infinity* of $\mathbb{H}^3$ has the following natural, intrinsic definition. Denote by $\mathcal{S}$ the set of all geodesic rays in $\mathbb{H}^3$ parameterized by arc-length on $[0, \infty)$, and define two elements of $\mathcal{S}$ to be equivalent if they do not diverge as their lengths go to infinity, in other words, define an equivalence relation R on $\mathcal{S}$ such that

$$\gamma_1 R \gamma_2 \iff \sup_{t \geq 0} d(\gamma_1(t), \gamma_2(t)) < \infty$$

for $\gamma_1, \gamma_2 \in \mathcal{S}$. Define $\partial\mathbb{H}^3 = \mathcal{S}/R$, $\overline{\mathbb{H}}^3 = \mathbb{H}^3 \cup \partial\mathbb{H}^3$ (as sets). We define a topology on $\overline{\mathbb{H}}^3$ so that $\mathbb{H}^3$ is open and inherits its own topology, and a basis of neighborhoods around $X \in \partial\mathbb{H}^3$ is defined as follows: choose γ in the class of X, let $x \in B^3$ be its starting point, let V be a neighborhood of $\dot{\gamma}(0)$ in the unit sphere of $T_x\mathbb{H}^3$, and let $r > 0$. Then set

$$U(\gamma, V, r) = \{\alpha(t) : \alpha \in \mathcal{S}, \alpha(0) = x, \dot{\alpha}(0) \in V, t > r\} \bigcup \{\langle\alpha\rangle_R : \alpha \in \mathcal{S}, \alpha(0) = x, \dot{\alpha}(0) \in V\},$$

where $\langle\alpha\rangle_R$ indicates the class of α in $\mathcal{S}/R$. Our basis comes from varying γ, V, and r. Considering the Poincaré ball model, we see that $\partial\mathbb{H}^3$ is canonically identified with $S^2 = \partial B^3$. Indeed, by construction, every geodesic γ corresponds to two points in $\partial\mathbb{H}^3$ (in B^3, these are the points where γ intersects ∂B^3); these are called the *ends* of γ. Conversely, given $p, q \in \partial\mathbb{H}^3$ with $p \neq q$, there is one and only one geodesic in $\mathbb{H}^3$ whose ends are p and q. Finally, note that $\text{Isom}^+(\mathbb{H}^3)$ permutes the geodesics of $\mathbb{H}^3$ and therefore extends to an action via Möbius transformations on $\partial\mathbb{H}^3$ in a canonical way. Thus we establish an intrinsic isomorphism between $\text{Isom}^+(\mathbb{H}^3)$ and $PSL_2(\mathbb{C})$.

6.2 Ideal Tetrahedra

In this section we discuss the properties of certain hyperbolic tetrahedra, which will play a central role when we begin to construct hyperbolic 3-manifolds.

Let Δ denote the standard oriented 3-simplex, and suppose we have an (orientation preserving) embedding $i : \Delta \hookrightarrow \mathbb{H}^3$. We say that $i(\Delta)$ is a *hyperbolic tetrahedron* if the image of each face of

Δ is a subset of a hyperbolic plane (it follows that the edges of Δ are sent to hyperbolic arcs). Suppose now that the image of i is actually $\overline{\mathbb{H}}^3$, but that only the 0-skeleton (the vertices) of Δ are allowed to land in $\partial\mathbb{H}^3$. We call each vertex of Δ which is on the sphere at infinity an *ideal vertex*; if all three vertices are ideal, we say that Δ is an *ideal hyperbolic tetrahedron*, or ideal tetrahedron for short. These are the tetrahedra we're interested in. In what follows, we will abuse notation and alternatively consider Δ as an abstract 3-simplex and as a subspace of $\overline{\mathbb{H}}^3$.

To investigate the properties of ideal tetrahedra, it is convenient to pass to another model of $\mathbb{H}^3$, the upper half-space model. Let $e = (0, 0, 1) \in \mathbb{R}^3$, and define $T : S^3 \to S^3$ to be the Möbius transformation such that

$$T(p) = e + 2J(J(p) - e).$$

T is a diffeomorphism from $B^3 \subset S^3$ to the upper-half space $U^3 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_3 > 0\}$ (essentially, stereographic projection from ∂B^3 to ∂U^3). Our second model of $\mathbb{H}^3$ is defined to be U^3 with the metric pulled back under T; this metric comes out to be $ds_x^2 = \frac{|dx|^2}{|x_3|^2}$, as one can check.

Definition 6.4. The Poincaré upper-half space U^3 (upper-half space for short) is the open set $\{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_3 > 0\} \subset \mathbb{R}^3$ with the Riemannian metric

$$ds_x^2 = \frac{|dx|^2}{|x_3|^2}$$

at each point $x \in B^3$.

In the upper-half space model, it is no longer the case that the entire boundary $\partial\mathbb{H}^3$ can be represented within the natural ambient space, however, it is still quite easy to understand: it consists of the plane ∂U^3 union a point, which we call ∞ or "the point at infinity in the upper-half space model," and which, topologically, would be the point at infinity if we compactified $\mathbb{R}^3 \supset U^3$ to S^3. Indeed, geodesics in U^3 fall into two categories: semi-circles perpendicular to ∂U^3, having both their ends on ∂U^3, and Euclidean lines perpendicular to ∂U^3, which have one end at ∞. (See Figure 9 for a picture of an ideal hyperbolic tetrahdron in U^3, showing all of these structures). Just as in B^3, the group $\mathrm{Isom}^+(U^3)$ acts on $\partial\mathbb{H}^3 \cong \partial U^3 \cup \{\infty\}$ by Möbius transformation, and therefore, for any $X \in \partial\mathbb{H}^3$, there is an orientation-preserving isometry of U^3 whose extension to the boundary sends X to ∞. This will help us visualize many things.

To investigate the properties of ideal tetrahedra, we will need the concept of a "sphere" around a point on $\partial\mathbb{H}^3$, called a *horosphere*. To give the definition, note that the spheres around a point X in $\mathbb{H}^n$ (or any n-dimensional geometry) can be defined as connected $n - 1$ manifolds orthogonal to all geodesics through X. Likewise, given a point $X \in \partial\mathbb{H}^3$, we define a horosphere around X to be a connected surface which is orthogonal to all geodesics with an end at X. In B^3 a horosphere centered at X is a Euclidean spheres contained in B^3 except for the point X, where it is tangent. The convex interior of a horosphere is called a *horoball* centered at X. Note that translation along, e.g., the diameter of B^3 ending at X permutes the horospheres (and horoballs) centered at X, and therefore they are all congruent.

It is even easier to visualize horospheres in U^3. Suppose we have some point X on ∂U^3; we may apply an isometry so that $X \mapsto \infty$. Then a horosphere at X is a plane parallel to ∂U^3 (see Figure 9). From this picture, and the explicit form of the metric on U^3, it follows immediately that there is a canonical Euclidean structure on each horosphere.

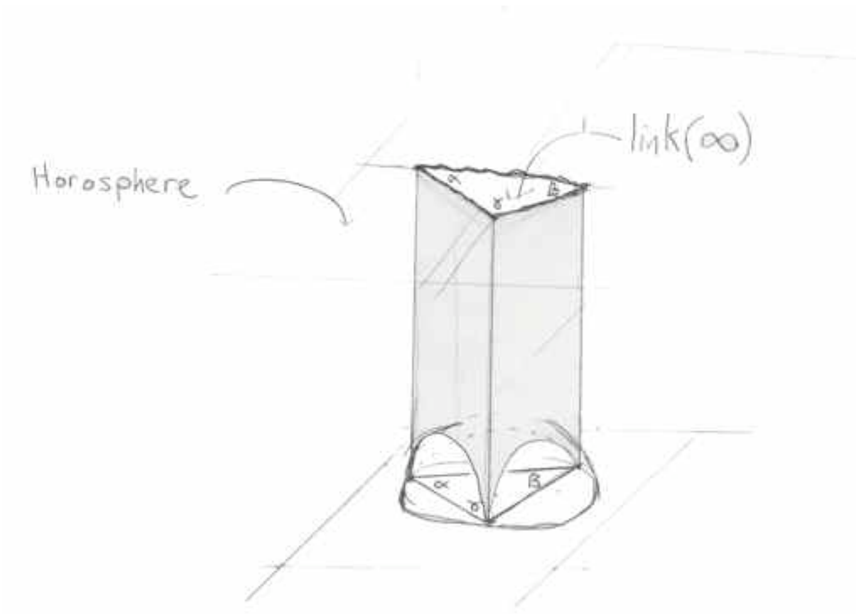

Figure 9: An ideal tetrahedron in U^3, intersecting a horosphere about ∞ (i.e, a Euclidean plane), to form its link $L(\infty)$, with dihedral angles α, β, γ

Remark 6.2. It is not at all obvious from looking at the standard horosphere in B^3 that it has a Euclidean structure; it appears to look like a sphere. One should think of the sphere's positive curvature as "canceling out" the negatively curved ambient hyperbolic space to arrive at a flat surface.

We will use the notion of horosphere to help us classify (oriented congruency classes of) ideal tetrahedra, proving that in fact, they have quite a simple moduli space. Specifically, let Δ be an ideal tetrahedron, and assume without loss of generality that one of its vertices, s, is at ∞. Then we can choose a sufficiently high horosphere H centered at s such that $\Delta \cap H$ is a Euclidean triangle, as in Figure 9. It is obvious that this triangle is well-defined up to Euclidean similarity; let $L(s)$ denote its similarity class, called the *link* of the vertex s.

We now note that the link $L(s)$ of an ideal tetrahedron, for any vertex, determines T up to congruence. To see this, suppose we have two ideal tetrahedra T, T' with vertices s, s' such that $L(s) = L(s')$. Use an isometry to move s to infinity; it follows, as in Figure 9, that the triangle in ∂U^3 formed by the three other vertices of T is in the similarity class $L(s)$. The same thing can be done for s' and T'. Then, note that given two similar triangles, in $\mathbb{R}^2 \cup \{\infty\}$, there is always a Möbius transformation sending one to the other and fixing ∞. The Poincaré extension of this Möbius transformation is an isometry of $\mathbb{H}^3$ sending the vertices of T to the vertices of T', and therefore sending T to T', as desired.

Note that the angles α, β, γ of $L(s)$ are precisely the so-called dihedral angles between adjacent faces of T at the corresponding edges, and it follows that the sum of the dihedral angles around a vertex of T is π. Denote the dihedral angles of all six edges of T by $\alpha, \beta, \gamma, \alpha', \beta', \gamma'$; we then have the equations

$$\alpha + \beta + \gamma = \alpha + \beta' + \gamma' = \alpha' + \beta + \gamma' = \alpha' + \beta' + \gamma = \pi$$

which reduces to the system

$$\begin{aligned}\alpha + \beta &= \alpha' + \beta' \\ \alpha + \gamma &= \alpha' + \gamma' \\ \beta + \gamma &= \beta' + \gamma'\end{aligned}$$

which has the unique solution $\alpha = \alpha'$, $\beta = \beta'$, $\gamma' = \gamma$. Therefore we actually conclude that the link of T does not depend on the choice of vertex (and furthermore, that the dihedral angles of opposite edges are equal).

We have established a bijection between congruence classes of ideal tetrahedra and similarity classes of triangles in E^2. We extend this to a bijection between oriented tetrahedra and oriented triangles in the natural way, by giving the link of T the orientation induced from T. Now we will classify oriented Euclidean triangles up to similarity. Simply choose a vertex v of $L(s)$ (corresponding to a choice of edge of T), label the other two vertices of $L(s)$ by u and t, so that v, t, u go clockwise around the origin if T is negatively oriented, and counter-clockwise if T is positively oriented; make the identification $\mathbb{R}^2 = \mathbb{C}$, and define

$$z(v) = \frac{t - v}{u - v} \in \mathbb{C}.$$

(Geometrically, this is the same thing as applying the unique orientation preserving similarity taking v to 0 and the v to either 1 or -1, depending on orientation). Note that im $(z(v)) > 0$ if T is positively oriented, < 0 if it is negatively oriented, and $= 0$ if T is degenerate, meaning all four of its vertices were on the boundary of some hyperbolic plane in $\partial\mathbb{H}^3$. See Figure 10.

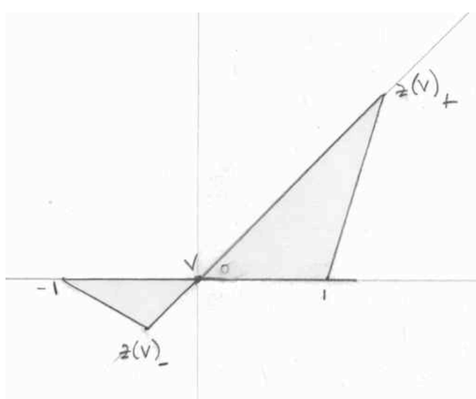

Figure 10: $z_+(v)$ is for the positively oriented case, $z_-(v)$ the negatively oriented case.

The other vertices have invariants $\frac{z(v)-1}{z(v)}$, $\frac{1}{1-z(v)}$

Remark 6.3. An ideal hyperbolic tetrahedron determines 4 points on S^2, which have a cross-ratio z, well defined up to $z \mapsto (z-1)/z \mapsto 1/(1-z)$. This agrees with $z(v)$ in the positively-oriented case.

Remark 6.4. There is a map from the set of (congruency classes of) oriented tetrahedra to itself, corresponding to switching the orientation; the corresponding map on the moduli space $\mathbb{C} - \{0,1\}$ is just inversion.

Remark 6.5. It follows directly from our construction that $\arg(z(v))$ is the dihedral angle at the corresponding edge.

Remark 6.6. Depending on the situation, we will write the modulus z as a function of tetrahedra, or as a function of edges, having fixed a particular tetrahedron. The context will make our usage clear.

6.3 Volume of Ideal Hyperbolic Tetrahedra

Another great thing about ideal hyperbolic tetrahedra is that, though they are not compact, they have finite volume. Indeed, define the *Lobachevsky function* $\Lambda(\theta)$, $\theta \in [-\pi, \pi]$, as

$$\Lambda(\theta) = -\int_0^\theta \log|2\sin u|du$$

Then it is well known, for an ideal hyperbolic tetrahedron Δ with dihedral angles α, β, and γ, that

Theorem 6.2. *Vol*$(\Delta) = \Lambda(\alpha) + \Lambda(\beta) + \Lambda(\gamma)$

See any of the references at the beginning of the section for a discussion.

Furthermore, define the *dilogarithm function*,

$$\mathrm{Li}_2(z) = \sum_{n=1}^{\infty} \frac{z^n}{n^2}, \quad \text{for } |z| \leq 1$$

analytically continued as

$$\mathrm{Li}_2(z) = -\int_0^z \frac{\log(1-w)}{w} dw$$

Letting z be a modulus for Δ, we have

Theorem 6.3. *Sign*$(Im(z)) \cdot Vol(\Delta) = D(z) = \arg(1-z)\log|z| - \mathrm{im}\,(Li_2(z))$

which we think of as the "algebraic volume" of Δ (it's positive if Δ is oriented positively, negative if Δ is oriented negatively, 0 if Δ is degenerate). $D(z)$ is the so-called *Bloch-Wigner* function.

6.4 Hyperbolic Manifolds

For some of the proofs we omit in this section, see [1, 23].

A *hyperbolic 3-manifold* is a smooth 3-manifold which is locally modeled on hyperbolic 3-space. More generally, suppose we have a smooth manifold X and a group G acting on X by diffeomorphism. Then an (X,G)-manifold is a smooth manifold made from patches of X which are glued together by restrictions of G. Formally, we have

Definition 6.5 ((X,G)-manifold). A smooth manifold M is said to be endowed with an (X,G)-*structure* if there is an open cover $\{U_i\}$ of M and a set of smooth open maps $\{\varphi_i\}$, $\varphi_i : U_i \to X$ such that

1. $\varphi_i : U_i \to \varphi_i(U_i)$ is a diffeomorphism
2. If $U_i \cap U_j \neq \emptyset$, then the restriction of $\varphi_i \circ \varphi_j^{-1} : X \to X$ to each connected component of $U_i \cap U_j$ agrees with the restriction of an element of G.

M is called an (X, G)-manifold.

We will make use of two particular sorts of (X, G) manifolds. In both cases, X will be a Riemannian manifold, with G acting on X either by isometries or similarities. In the former case, M inherits its own Riemannian structure, and is called a *geometric* manifold. In the second, it is called a *similarity* manifold. (This latter concept can seem a little unfamiliar. However, note that we've already encountered an example: the fact that the link $L(v)$ of an ideal tetrahedron has a Euclidean structure, defined up to similarity, naturally gives it a $(E, \mathrm{Simm(E)})$ structure). In both of these cases, the action of G on X is *rigid*, meaning that each element $g \in G$ is uniquely determined by its action on any open subset of X. When G acts on X rigidly, we can construct a "developing" map $D : \widetilde{M} \to X$ ($\widetilde{M}$ is the universal cover of M) which naturally induces the (X, G)-structure on $\widetilde{M}$, and therefore on M. This gives us a way to "develop," or unravel, simply connected pieces of M into X. This will often be helpful, as we usually have a much better understanding of the geometry of X, than of the geometry of M.

Theorem 6.4. *Let M be an (X, G)-manifold, and suppose the action of G on X is rigid. Then there exists a map $D : \widetilde{M} \to X$, which is an immersion inducing the (X, G) structure on $\widetilde{M}$ (which is itself induced by the structure on M).*

To construct D, we start with any chart $\phi : U \to X$ of the (X, G) structure on M. Next, choose any path $\alpha \subset M$; it is not too hard to see, using the rigidity of the action of G, that we can analytically continue ϕ along α. Thus, considering all paths in M, we obtain a map from the universal cover of M into X. Starting with a different chart simply changes D by post-composition with an element of G, indeed, D is unique up to composition with elements of G.

The monodromy of M is a map $\pi_1(M) \to \mathrm{Deck}(\widetilde{M})$, defined up to conjugation. Now, by the uniqueness of D, we have $T \circ D = \gamma \circ D$ for $T \in \mathrm{Deck}(\widetilde{M})$ and some unique $\gamma \in G$. The correspondence $T \mapsto \gamma$ actually defines homomorphism $\mathrm{Deck}(\widetilde{M}) \to G$, and by first applying the monodromy map, we get a map $H : \pi_1(M) \to G$ called the *holonomy* of M, well-defined up to conjugation.

The holonomy gives us some information about "sub-structures." Indeed, suppose that the (X, G) structure on M actually restricts to an (X, K) structure, where $K \subset G$ is some subgroup (so also acting rigidly). Then the construction of a holonomy for the (X, K)-structure gives a holonomy for the (X, G) structure at the same time, so it follows that:

Theorem 6.5. *Suppose $K \subset G$ is a subgroup and M is an (X, G) manifold. Then the (X, G) structure on M actually restricts to an (X, K) structure only if $H(\pi_1(M)) \subset K$ for some holonomy H.*

(Of course, if K is a normal subgroup, then "a holonomy" can be replaced by "all holonomies"). The converse of Theorem 6.5 also holds, though we won't need it.

What's more, there is a special case in which the holonomy determines the entire (X, G) structure. If D is surjective, then it is a covering map, and M is said to be *complete*. If X is also simply connected, it follows immediately that D is in fact a homeomorphism, so $\widetilde{M} \cong X$ and $M = X/H(\pi_1(M))$, and we see that in this case, the holonomy determines the (X, G) structure on M. Now, suppose M is a geometric manifold, so it has a metric induced from the Riemannian manifold X. Then we have

Theorem 6.6. *M is complete in the above sense if and only if it is complete as a metric space with the induced metric.*

For the proof, see [23].

We now rigorously define an (oriented) hyperbolic 3-manifold: a smooth 3-manifold M with a $(\mathbb{H}^3, \mathrm{Isom}^+(\mathbb{H}^3))$ structure. Furthermore, as already mentioned, a Euclidean similarity surface is, by definition, a surface M with a $(E, \mathrm{Simm}(E))$ structure. By Theorem 6.6 and the proceeding discussion, a hyperbolic 3-manifold is complete if and only if it is can be realized as $\mathbb{H}^3/\Gamma$, where $\Gamma \subset \mathrm{Isom}^+(\mathbb{H}^3)$ is a discrete subgroup of hyperbolic isometries, acting properly discontinuously on $\mathbb{H}^3$.

Remark 6.7. Note that it is a famous property of 3 (and higher) dimensional *complete* hyperbolic manfiolds that they are *rigid*; this is called Mostow Rigidity. In other words, the hyperbolic structure on such a manifold is actually a *topological* invariant of the underlying manifold. For example, if the complement of a knot has a complete hyperbolic structure with finite volume, this volume is an invariant of the knot.

6.5 Gluing Tetrahedra

Suppose we have a finite set $\{\Delta_i\}$ of oriented (topological) tetrahedra, and a pairing between their faces which reverses orientation (and doesn't send any face to itself). This pairing induces an equivalence relation R on $\coprod \Delta_i$ such that $\left(\coprod \Delta_i\right)/R$ is a closed, oriented, triangulated 3-manifold $\bar{M}$, except possibly in a neighborhood of each coset of vertices. Call each such coset a *cusp* of $\bar{M}$; for each cusp $v \in \bar{M}$, we can choose a regular neighborhood $L(v)$, disjoint from all other cusps of $\bar{M}$, such that $L(v) - \{v\}$ is homeomorphic to $V \times (0, \infty)$ for some closed orientable surface V. Define the orientable, non-compact manifold M to be $\bar{M}$ with its cusps removed. We still have $L(v) \subset M$ for each cusp; we continue to refer to v as a cusp of M (it is obvious that, fixing M, any other triangulation of $\bar{M}$ corresponding to M must also give the same set of cusps). We call $L(v)$ the *link* of the cusp v.

Remark 6.8. There is an intrinsic definition of cusp which applies to general 3-manifolds, but we will not go down that route.

Note that the triangulation of $\bar{M}$ gives a triangulation of M, except that it is now by topological *ideal* tetrahedra, that is, 3-simplices with their vertices removed. Let T denote the data of this triangulation, so that we write $\Delta \in T$ to indicate that Δ is one of the (now ideal) tetrahedra used in triangulating M. In this section, the pair (M, T) will indicate a 3-manifold M with such a *topological ideal triangulation* T (non-empty, so that M is *a priori* non-compact).

Take such a pair (M, T), and suppose that $M = \mathbb{H}^3/\Gamma$ is actually a complete, finite-volume hyperbolic manifold, so that, by the results of the previous section, we can choose a covering map $\pi : \mathbb{H}^3 \to M$. Our goal is to try to "straighten out" the ideal tetrahedra from T into a *hyperbolic* ideal triangulation of M, still called T. More formally, we will try to put a hyperbolic structure on each $\Delta \in T$, such that $\pi^{-1}(\Delta)$ is an isometry (for each choice of lift); when this can be done, it can be done uniquely. Then the hyperbolic ideal tetrahedra from T induce the hyperbolic structure of M (note that this is *not* the same as saying that these hyperbolic tetrahedra can be glued up to give a manifold isometric to M, as we shall see). What's more, we've already seen that choosing a hyperbolic structure on an ideal tetrahedron Δ amounts to choosing an edge of Δ and a modulus from $\mathbb{C}-\{0,1\}$; we will describe algebraic equations which the moduli of these hyperbolic tetrahedra must automatically satisfy, called the *hyperbolic gluing equations* of T. Finally, we can use these moduli to compute the volume of M.

Remark 6.9. Doing this process in reverse, that is, starting with a 3-manifold M *not* known to be hyperbolic, triangulating it topologically, and then seeing if the triangulation can be used to give M a hyperbolic structure, is more subtle, as we shall see, though of course it is much more interesting, if one wants to find new hyperbolic 3-manifolds.

We start with a well-known comment on isometries of $\mathbb{H}^3$. They can be divided into three categories: the first, of *elliptic* type, have fixed points in $\mathbb{H}^3$. Otherwise, *parabolic* isometries are, up to conjugation, Poincaré extensions of translations of $\partial\mathbb{H}^3$, and therefore fix one point on the sphere at infinity; it can be shown, furthermore, that they set-wise fix each horosphere centered at this fixed point, (and therefore, up to conjugacy, restrict to Euclidean translations on the horospheres), see [1]. *Loxodromic* isometries are, up to conjugacy, extensions of dilations and fix two points of the sphere at infinity. Recalling that $M = \mathbb{H}^3/\Gamma$, it follows that no elements of Γ are elliptic. One can also show that Γ contains parabolic elements if and only if M is non-complete. More specifically, let $L(v)$ be a link of M and let V be the surface that corresponds to taking a slice of $L(v)$; then $\pi_1(V)$ is naturally realized as a subgroup of $\pi_1(M)$, and one can show that, under any holonomy, its image must consist of parabolic elements fixing the same point $w \in \partial\mathbb{H}^3$ (this point "covers" the cusp v) (this group of parabolic isometries will be referred to as the holonomy *of the link* $L(v)$).

Now, consider $\Delta \in T$ as a simply connected subspace of M. We have

Theorem 6.7. *The closure in* $\bar{\mathbb{H}}^3$ *of any connected component of* $\pi^{-1}(\Delta)$ *contains at most 4 points.*

Proof. Suppose $\{x_n\}$ is a sequence of points in Δ, converging to a vertex v in the closure $\bar{\Delta}$. Then the lift of this sequence is a divergent sequence in $\mathbb{H}^3$, and we can choose a subsequence which converges to a point $x \in \partial\mathbb{H}^3$, which, by the previous discussion, is fixed by the holonomy of the link $L(v)$. The lift of any regular neighborhood of v in M must be stabilized by this holonomy, and therefore must be contained in some horoball centered at x. Therefore the sequence $\{x_n\}$ can have at most one limit point in $\partial\mathbb{H}^3$, as desired. □

We will obtain less than 4 points from the above construction if and only one or more edges of Δ lifts to a loop in $\bar{\mathbb{H}}^3$. We call such an edge *homotopically trivial.* Suppose for now that none of the edges of T are homotopically trivial (later, will discuss conditions on the topology of (M,T) for this to be the case). Because 4 distinct points on $\partial\mathbb{H}^3$ uniquely determine a hyperbolic ideal tetrahedron, we can, by Theorem 6.7, associate a hyperbolic tetrahedron to each $\Delta \in T$, well-defined up to congruence, on which, by construction, π^{-1} restricts to an isometry. We have seen that every hyperbolic tetrahedron is determined by an element of $\mathbb{C} - \{0,1\}$, once we choose an edge. Fix a generic $\Delta \in T$, and write $z(\Delta)$ for one of the moduli.

Remark that our construction of moduli for tetrahedra depended on a choice of orientation. Now, each topological ideal tetrahedron $\Delta \in T$ is positively oriented, by construction (the orientation of M is chosen to be the one induced by the triangulation), and π is orientation preserving.

However, recall that we are *not* lifting the topological tetrahedra by π^{-1}; instead, we lift just the vertices, and take the tetrahedron they determine. Now, the vertices, ordered by the orientation on Δ, are enough to induce an orientation on this lifted tetrahedron, however, this orientation may now be negative (with respect to the fixed orientation of $\mathbb{H}^3$) instead of positive. Geometrically, this corresponds to the following situation, where a face of the topological lifting is essentially on the "wrong side" of the hyperbolic tetrahedron:

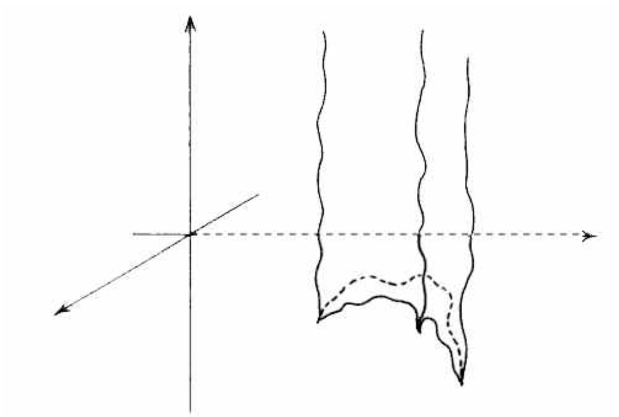

Figure 11: A positively oriented topological tetrahedron (pictured in U^3) whose straightening will be negatively oriented

By construction, $z(\Delta)$ is in the lower half plane if and only if such a "folding over" occurs (because this is the only situation where the choice of hyperbolic tetrahedron to replace Δ will be negatively oriented). When such a phenomenon occurs, the hyperbolic triangulation of M by our straightened tetrahedra will not be embedded.

Furthermore, regardless of orientation, the four lifted vertices of Δ may actually be on a hyperplane, in which case the hyperbolic tetrahedron that we get is degenerate, and its moduli are on the real line.

For these two reasons, the induced hyperbolic triangulation T cannot actually be glued up via hyperbolic isometries to induce a hyperbolic structure on M; this is our follow-up to Remark 6.8. However, the geometric picture still ensures that the tetrahedra fit together around each edge (despite the fact that they may overlap). In other words, choose an edge in the hyperbolic triangulation, let $\{E_i\}_1^m$ be the set of tetrahedra from T which abut this edge, and let $z(E_i)$ be the modulus of E_i corresponding to the chosen edge. Then we have

$$\prod_{i=1}^{m} z(E_i) = 1 \tag{28}$$

$$\sum_{i=1}^{m} \arg(z(E_i)) = 2\pi, \tag{29}$$

recalling that the modulus at a vertex is, by definition, the ratio of the other two edges. These are called the *edge equations* of the triangulation T. Here's a picture in the positively oriented case:

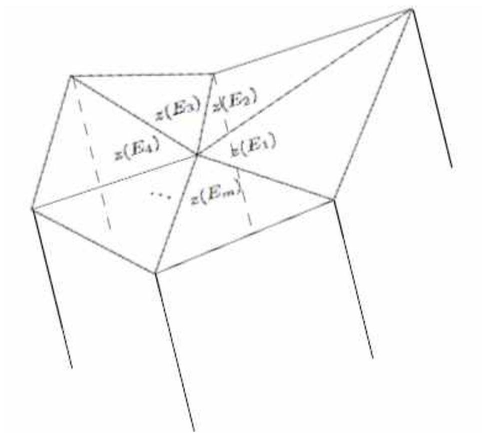

Figure 12: Links of tetrahedra fitting around an edge

Furthermore, the moduli satisfy equations corresponding to the fact that the hyperbolic structure on M is complete. Specifically, choose a cusp $v \in M$, with link $L(v) \cong V \times (0, \infty)$; in abuse of notation, we let $L(v)$ denote a particular slice of the link, i.e., a copy of V in M. (This is standard practice). First remark that the triangulation T induces a triangulation T' of $L(v)$ (possibly degenerate, e.g., non-embedded). (It also induces an orientation on V, proving that $L(v)$ is orientable). Furthermore, once we have chosen a hyperbolic structure on the tetrahedra in T, the triangles in T' correspond to links of those tetrahedra, so they each have an induced Euclidean similarity structure. What's more, by the edge equations, these links fit together geometrically along their edges, so the similarity structure globalizes to all of $L(v)$.

Remark 6.10. Note that the only orientable surface which can be given a Euclidean similarity structure is the torus. So (assuming that all non-compact hyperbolic 3-manifolds with finite volume can be ideally topologically triangulated, which is true), we have shown *a priori* that the links of the cusps of all finite-volume hyperbolic 3-manifolds are tori.

Theorem 6.8. *If the manifold M is complete, then the similarity structure on $L(v)$, for each cusp $v \in M$, is actually a geometric structure.*

Proof. In what follows, by simplicial loop, we mean a simple loop on a triangulated surface which consists of edges of the triangulation.

Assume M is complete, fix a covering map $\pi : \mathbb{H}^3 \to M$, and let $L(v)$ be any of the triangulated links of M. We can remove two simplicial loops on $L(v)$ which generate its fundamental group, obtaining a simply connected subspace of $L(v)$ which we can develop into U^3 via π, leaving us with a horizontal subset of U^3 whose closure is a triangulated polygon. The identifications on the edges of this polygon which glue it up into $L(v)$ are induced by isometries belonging to a discrete subgroup $\Gamma \in \mathrm{Isom}^+(\mathbb{H}^3)$, which are parabolic fixing the same point $v \in \partial U^3$, which we can assume is ∞. Therefore they act Euclidean isometries along a horosphere, inducing a Euclidean structure on $L(v)$. $\square$

In particular, the holonomy of $L(v)$ is contained in $\mathrm{Isom}^+(E^2)$ by Theorem 6.5, and therefore, thinking of similarities of $E^2 \cong \mathbb{C}$ as maps $ax + b$, $b, x \in \mathbb{C}, a \in \mathbb{C}^*$, it follows that the images of the generators of $\pi_1(L(v)) = \mathbb{Z}^2$ under any holonomy have derivative one. (Note that the derivative of a similarity is conjugation-invariant). Now, suppose we have some directed simplicial loop γ in $L(v)$. Fixing a side of γ (say, "on the right" as we move along it), we record, at each vertex in γ, a number of moduli of tetrahedra. Then, it is not hard to see that the product of all these moduli is precisely the derivative of the holonomy of γ (recall again that the moduli are just ratios of adjacent sides, and, for example, construction from the proof of Theorem 6.8). Thus, for every oriented simplicial loop in T, we get an equation, since this product must equal 1 if M is to be complete. These are called the *cusp equations* of T.

Finally, we remark that though the hyperbolic triangulation of M that we have constructed may not be embedded, the homology class (i.e., the fundamental class of M) represented by the two sets of simplices (hyperbolic and topological) agree (the argument is straightforward, see [1, p. 106] for details). Therefore the map from the hyperbolic triangulation into M has degree one, and it follows that if we sum over the algebraic volume

$$\sum_{z(\Delta)|\Delta\in T} D(z)$$

(implicitly making some choice of modulus, for each $\Delta \in T$), we get the volume of M.

For convenience, given an ideal topological triangulation T of a 3-manifold M, we refer to cusp and edge equations together as the *hyperbolic gluing equations* of T.

Remark 6.11. Except for this issue of straightenings which are degenerate or folded-over, it can be shown that the hyperbolic gluing equations are actually sufficient, not only necessary, conditions on the moduli of a hyperbolic ideal triangulation. In other words, if we have a topological ideal triangulation (M, T) of some 3-manifold M, we can write down the same set of equations (even though there is no hyperbolic structure *a priori*, as we were assuming before; we can still write down the equations, formally); then, if we can find a solution such that the moduli live in the upper half plane, this gives a unique complete hyperbolic structure to M. Thus we can think of the hyperbolic gluing equations as essentially determining the hyperbolic structure.

7 The Volume Conjecture

Let $L \subset S^3$ be a knot such that $M = S^3 - L$ is a complete hyperbolic 3-manifold; L is called a *hyperbolic knot*. From our exposition so far, hyperbolic geometry and quantum knot invariants, though they both describe properties of 3-dimensional situations, seem to have absolutely no connection. Thus the following conjecture is quite remarkable:

Conjecture 7.1 (Hyperbolic Volume Conjecture).

$$2\pi \times \lim_{N\to\infty} \frac{|\log F'_n(L)|}{N} = vol(M)$$

In words, the asymptotic growth of the colored Jones polynomial determines the hyperbolic volume of the knot complement.

In this conjecture, as in the second part of §5, we are implicitly setting $q = e^{2\pi i/N}$ in $F'_N(L)$. We fix this convention throughout.

In this section, we will give an account of the connection between these two invariants of L, and outline an incomplete, but quite tantalizing, framework in which the Volume Conjecture might be proven. We follow [29]; as far as the author is aware, this should be called Yokota's (pseudo)proof. In particular, we give a proof of the following theorem, whose precise statement must wait until the end of the section:

Theorem 7.1. *For large N, $F_n'(L)$ can be approximated as*

$$F_N'(L) \approx \sum_{s_1, s_2, \ldots, s_n} \chi(q^{s_1}, q^{s_2}, \ldots, q^{s_m}) \cdot e^{\frac{N}{2\pi i} V(q^{s_1}, q^{s_2}, \ldots, q^{s_m})}, \quad q = e^{\frac{2\pi i}{N}} \tag{30}$$

where $V(z_1, z_2, \ldots z_m) \in \mathbb{C}$ is a certain sum of dilogarithm functions, and the absolute value of $\chi(z_1, z_2, \ldots, z_m)$ grows at most polynomially as $N \to \infty$, at least when $z_1, \ldots z_m$ are on the unit circle (as in (30)*). (The sum in* (30) *is over certain integers between* 1 *and N, which will be explicated in what follows). Furthermore, there is a branch V_0 of V such that a certain solution $\zeta = (\zeta_1, \ldots, \zeta_m)$ to the stationary phase equations*

$$\frac{\partial V_0}{\partial z_1} = 0, \quad \frac{\partial V_0}{\partial z_2} = 0, \cdots, \frac{\partial V_0}{\partial z_m} = 0$$

corresponds to a solution to the hyperbolic gluing equations for a certain triangulation of M, and we have

$$\mathit{vol}(M) = \mathit{Im}(V(\zeta_1, \zeta_2, \ldots, \zeta_m))$$

The existence of such a potential function for a set of hyperbolic gluing equations, whose imaginary part gives the hyperbolic volume, is an interesting result in itself, investigated for the first time, I believe, by Neumann and Zagier in [19]. The results of this section give a combinatorial way to deduce this function, at least for a particular set of triangulated hyperbolic 3-manifolds.

To see how Theorem 7.1 relates to Conjecture 1, note that, heuristically, we might hope to be able to replace the sum in (30) by an integral, as N goes to ∞; then (30) becomes an oscillatory integral. Then by the so-called saddle-point or stationary-phase approximation, the integral would converge to $e^{(N/2\pi i)V_i(\alpha)}$, α being the critical point of V or one of its branches such that $V(\alpha)$ has maximal imaginary part. Conjecturally, the ζ in Theorem 7.1 satisfies this property. Then we would have $2\pi \times \lim_{N\to\infty} \frac{|\log F_n'(L)|}{N} = \mathrm{Im}(V(\zeta)) = \mathrm{vol}(M)$, as desired. Note, however, that ζ may not be on the unit circle, whereas in (30), V is only evaluated at points on the unit circle. And once we move off the unit circle, the function χ might grow exponentially (and even if it does not, we don't know if we can deform the contour of integration off of the unit circle, while preserving the asymptotics). Thus the Volume Conjecture is reduced to a number of subtle analytic questions, and it is unclear how these relate to the general geometric picture which we here describe.

It may be helpful to give a brief overview of what follows. We start out with a diagram D for our knot L, and from this diagram, we obtain an ideal topological triangulation of $M = S^3 - L$ with two points removed. We collapse these two points into the knot, degenerating many of the original tetrahedra, and obtaining an ideal topological triangulation of M. We then describe the edges of this triangulation $\mathcal{S}$ explicitly, to uncover the hyperbolic gluing equations on the moduli of the tetrahedra.

The first serendipitous result is that the hyperbolic gluing equations can be rewritten to depend not on a function from the set of tetrahedra of $\mathcal{S}$ to $\mathbb{C}$, but a function from the edges of the diagram to $\mathbb{C}$. This should be seen as quite an important step, in view of the fact that our formalism for computing $F'_{N-1}(L)$ is in terms of states on the edges of the knot diagram. Next, we define a function $V(z)$ which is the sum of one dilogarithm per tetrahedra in $\mathcal{S}$. Using some of the remaining hyperbolic gluing equations, we show that this function has a critical point at the solution to these equations, and that at this critical point, its imaginary part gives the hyperbolic volume.

To finish, we compute $F'_{N-1}(L)$ from the knot diagram. There is a close connection between the asymptotics of the quantum factorials appearing in the R-matrix for $F'_{N-1}(L)$, and the dilogarithm function. However, it seems at first that $F'_{N-1}(L)$ will give the volume of the original triangulation of M missing two points, before we collapsed it to a true triangulation. On closer inspection, however, a series of miraculous cancellations occur, and we eliminate all factorials from $F'_{N-1}(L)$ corresponding to the tetrahedra that have collapsed. Theorem 7.1 follows immediately.

7.1 Preliminaries

Let L be a knot in S^3, and let D be a diagram of L, which we will think of as a 4-valent graph lying in S^2. In this section, we prepare some notation concerning D.

Let $n \in \mathbb{N}$ be the number of crossings of D, and let $\{X_1, \ldots X_n\}$ denote these crossings, which are the vertices of D; we will use the terms crossing and vertex interchangeably.

Theorem 7.2. *D has $n+2$ faces*

Proof. It can be easily checked that D has $2n$ edges, so the result follows from $\chi(S^2) = 2$. □

Let $\{R_0, \ldots R_{n+1}\}$ denote the faces of D. We choose a non-crossing point $p \in D$ to be the point at infinity of $S^2 = \mathbb{R}^2 \cup \infty$; our heuristic will be to consider D "broken" at p, to obtain a (1,1) tangle, though we will not incorporate this "breaking" into the diagram D (i.e., we merely consider p as a distinguished point on D). Without loss of generality, we take R_0 and R_{n+1} to be the two faces of D which have p on their boundary (of course, we can assume they're distinct, or else we can reduce the number of crossings in D). In our illustrations, we will depict D in $\mathbb{R}^2$, with the edge intersecting p broken in half, so that R_0 and R_{n+1} are the two unbounded faces. See Figure 13 (though we have not yet described all the structure from the figure).

Let D^* denote the dual graph of D, and write $\{X_1^*, \ldots X_n^*\}$ for the faces of D^*, and $\{R_0^*, \ldots R_{n+1}^*\}$ for its vertices. Note that $D \cap D^*$ consists of one point for each edge of D; we call this point the *midpoint* of the corresponding edge. An overpass of D is the obvious thing: a maximal continuous arc of the diagram, going from the midpoint of some edge to the midpoint of another edge, which only crosses *over* all the vertices it contains. We define underpass analogously, and refer to them collectively as *passes*. Let $\mathcal{P}_D$ denote the set of passes of D , $\mathcal{P}_D^+$ the set of overpasses, and $\mathcal{P}_D^-$ the set of underpasses (we will be switching to a different diagram later on, so it is important to keep track of D in our notation). Let $\mathcal{E}_D$ denote the edges of D. An edge of D is called *alternating* if it goes over one of its vertices and under the other; an alternating edge intersects an overpass and an underpass. Otherwise, the edge is called non-alternating, and it is contained entirely in a single pass. Let $\mathcal{K}_D$ denote the set of non-alternating edges of D, $\mathcal{K}_D^+$ the non-alternating edges contained in an overpass, and $\mathcal{K}_D^-$ the non-alternating edges contained in an underpass. Without loss of generality, we assume that p, the broken point, is the midpoint of an alternating edge. Suppose that φ_+ and φ_- are the underpass and overpass, respectively, which contain p. Then, start at p, continue along D in the direction of φ_+, and denote by X_a the first undercrossing you come to. Similarly, starting at p and continuing in the direction of φ_-, let X_b be the first overcrossing you

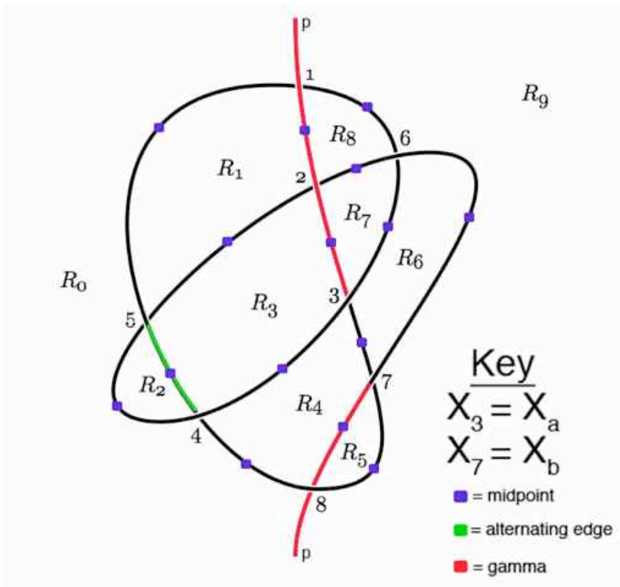

Figure 13: The knot 8_{21}, which we will use as our example. We've labeled an example of an alternating edge, as well as the midpoints, and the arc γ.

come to (see Figure 13). We also let X_1 and X_n be the vertices adjacent to p, with X_1 contained in the overpass and X_n in the underpass.

Remark 7.1. We think of all the objects defined in this section, e.g. crossings, edges, passes, faces, etc., as *closed* subsets of S^2, and use corresponding notation: e.g., for $\varphi \in \mathcal{P}_D$, $X_\nu \in \varphi$ means that the vertex X_ν is contained in the pass φ.

In the rest of this section, we cite two lemmas, which allow us to make certain assumptions on D, up to applying an ambient isotopy. For their proofs, see [29] (the arguments are generally straightforward: if such a thing were not so, one could reduce the crossing number of D). The first is

Lemma 7.1. *Without loss of generality, we may assume $X_a \neq X_b$*

Let γ be the arc of D connecting X_a to X_b and containing p. Again letting φ_+ and φ_- be the underpass and overpass containing p, Let $\mathcal{N}_1$ denote the set of crossings contained in φ_+, and $\mathcal{N}_0$ the set of crossings contained in φ_-. (It follows that $\mathcal{N}_1 \cup \{X_a, X_b\} \cup \mathcal{N}_0$ is the set of crossings contained in γ).

Suppose X_ν is a crossing of D, and R_μ is a face which abuts X_ν. Then there is a corresponding face $X_\nu^* \cap R_\mu$ of the graph $D \cup D^*$; we call this face an *angle* of X_ν, and we say that the pair (ν, μ) is an angle of D. Given a vertex X_ν, let $\mathcal{N}_\nu^*$ denote the set of faces abutting X_ν. Likewise, given a face R_μ, let $\mathcal{R}_\mu$ denote the vertices which border that face.

Lemma 7.2. *Without loss of generality, we can assume $\mathcal{N}_1^* \cap \mathcal{N}_n^* = \{0, n+1\}$. We can also assume that these (X_1 and X_n) are the only vertices in γ contained in $\mathcal{R}_0, \mathcal{R}_{n+1}$, and that the overpasses φ_+ and φ_- do not intersect each other. Finally, we may assume $|\mathcal{N}_\nu^*|=4$ for all crossings ν.*

For $\varphi \in \mathcal{P}_D$, let m_φ denote the midpoint of φ, which will be either a crossing or a midpoint of an edge of D, according as φ contains an odd or an even number of crossings, respectively. Likewise, for $\epsilon \in \mathcal{E}_D$, let m_ϵ denote the midpoint of ϵ.

7.2 Triangulation of $\dot{M} = S^3 - \{L \cup \{\pm\infty\}\}$

Now that we have worked out our notation, we introduce an explicit triangulation of M, following from the combinatorics of D. Such a triangulation is initially discussed in [22], and expounded on in [29], though note that quite a similar construction is considered as early as [27], and probably by many others with regards to the hyperbolic geometry of knot complements.

Fix two points $\pm\infty \in S^3$, assumed to be disjoint from L, which we think of as the poles of the 3-sphere. Define $\dot{M} = M - \{\pm\infty\}$. Thinking of $S^3 - \{\pm\infty\}$ as $S^2 \times (-\infty, \infty)$, we consider the diagram D as lying in $S^2 \times \{0\}$. Let $p : S^3 \to S^2$ be the projection defining D. Then we can speak of passes of L simply by lifting the passes of D via p^{-1}, in the obvious way. To help visualize $L \subset S^3$, we assume that "most" of each overpass of L lies in $S^2 \times \{1\}$, and "most" of each underpass lies in $S^2 \times \{-1\}$, with dips in between the passes. In fact, we can assume L intersects $S^2 \times \{0\}$ precisely in the lifts of the midpoints of the alternating edges of D (this is explicated in figures to come).

Now, place one ideal topological octahedron O_ν between the two lifts of each crossing X_ν of D, so that two of its (ideal) vertices are on L and project onto X_ν (see Figure 14); of the other vertices, two will project onto an overpass of D, and two onto an underpass of D. We drag the former to $-\infty$, and the latter to ∞, as shown, so none of the ideal vertices are in our manifold. To get an ideal triangulation of $\dot{M}$, we divide each of these octahedra into four tetrahedra; then there is one tetrahedron for each face of $D \cup D^*$, in particular, each tetrahedron projects onto a unique angle of D. We define $S_{\nu\mu}$ to be the tetrahedron projecting onto (ν, μ). Following along the knot, each face of each tetrahedron is paired with exactly one other face of some other tetrahedron (by Lemma 7.2, no two paired faces belong to the same tetrahedron). See Figure 15. Giving the tetrahedra the orientation induced from $\dot{M}$, we obtain an ideal topological triangulation, which we will write $(\dot{M}, \dot{\mathcal{S}})$.

It is not difficult to write down the edges of $\dot{\mathcal{S}}$. Indeed, they can be written as follows, where we have also recorded the parts of D that they are each naturally in bijection with (in particular, they each can be chosen to naturally project onto the corresponding structure under p. This should allow the reader to find them e.g. in Figure 15).

1. $\dot{E}_\nu = p^{-1}(X_\nu) \cap \{S^2 \times (-1, 1)\}$ (crossings)

2. $\dot{F}_\mu = p^{-1}(R^*_\mu)$ (faces)

3. $\dot{I}_\lambda = \begin{cases} p^{-1}(m_\varphi) \cap \{S^2 \times (1, \infty)\} \text{ if } \varphi \in \mathcal{P}^+_D \\ p^{-1}(m_\varphi) \cap \{S^2 \times (-\infty, -1)\} \text{ if } \varphi \in \mathcal{P}^-_D \end{cases}$ (midpoints of passes)

4. $\dot{H}_k = \begin{cases} p^{-1}(m_\kappa) \cap \{S^2 \times (-\infty, 1)\} \text{ if } \kappa \in \mathcal{K}_+ \\ p^{-1}(m_\kappa) \cap \{S^2 \times (-1, \infty)\} \text{ if } \kappa \in \mathcal{K}_- \end{cases}$ (midpoints of non-alternating edges)

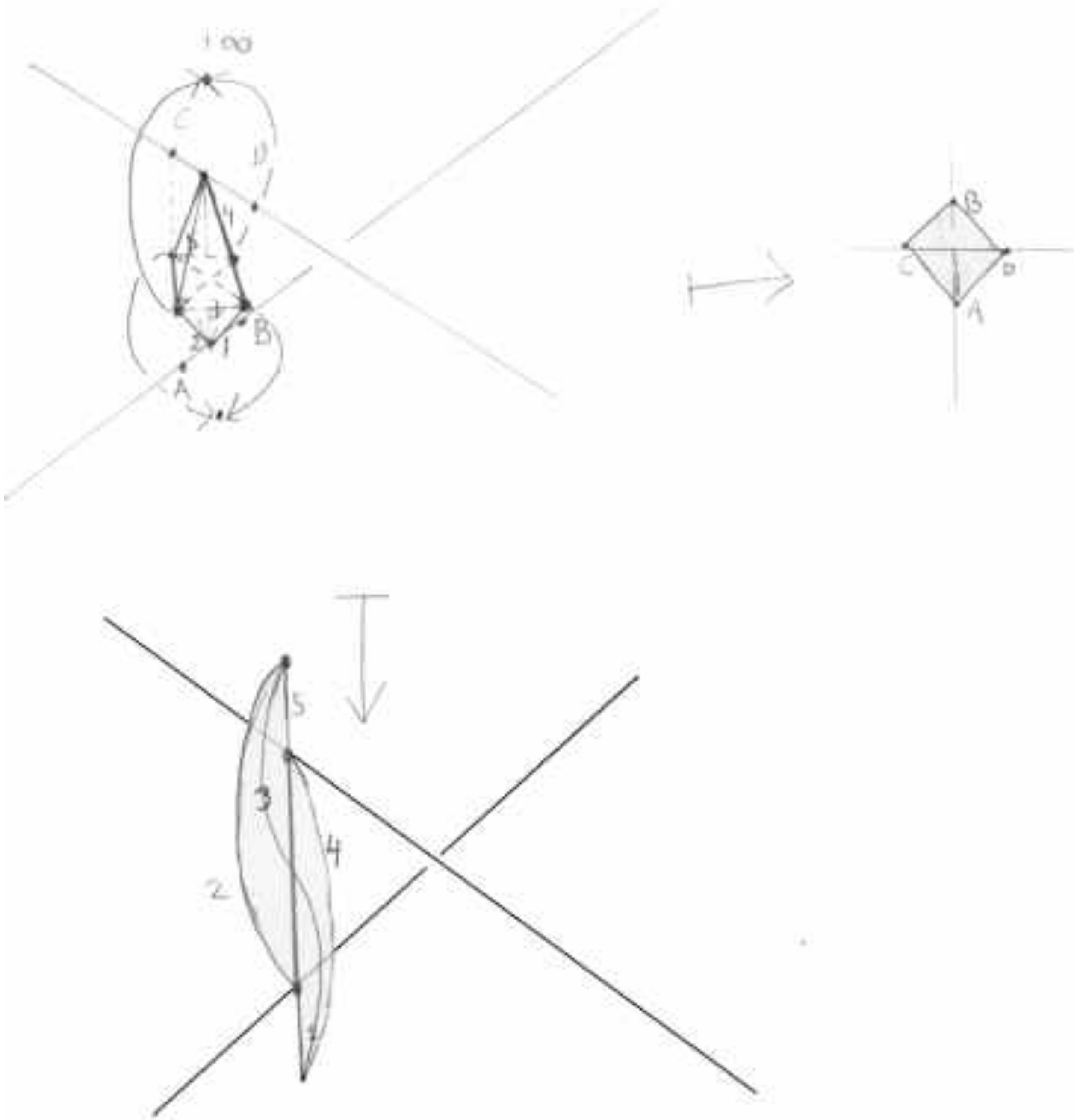

Figure 14: At top left, one of the octahedra O_ν. At top right, the projection of O_ν onto D *before* we've dragged the vertices. At bottom, one of the tetrahedra obtained from O_ν, after having its vertices pulled to $\pm\infty$.

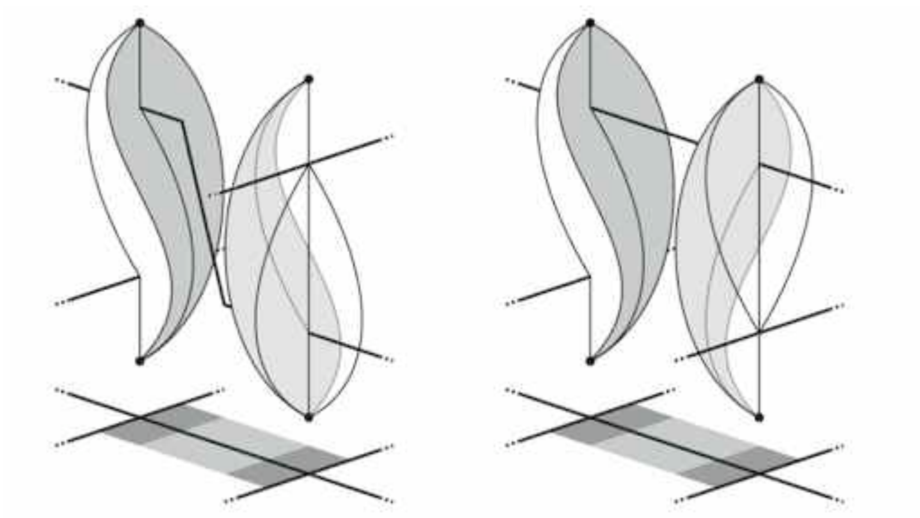

Figure 15:

7.3 Degeneration into a triangulation of $M = S^3 - L$

We now have an ideal triangulation $\dot{\mathcal{S}}$ of $\dot{M}$, from which we would like to produce an ideal triangulation $\mathcal{S}$ of M. Let $Q = O_1 \cap O_n$, which is a bigon in S^3 connecting $\pm\infty$ (see Figure 16). It is clear that $\dot{M} - Q$ and M are homeomorphic (we will discuss a particular homeomorphism later in this section). What's more, Q projects onto the broken point p, so we can think of ourselves as actually investigating the complement of a (1,1)-tangle rather than a knot, which perhaps motivates the presence of a (1,1)-tangle invariant in the Volume Conjecture.

In order to understand how to go from a triangulation of $\dot{M}$ to one of M, we go on a brief digression to discuss the linear degeneration of 3-simplices. Given any 3-simplex, and a choice of two of its vertices, there is a unique "linear degeneration" into a triangle, which identifies the two vertices. However, if the 3-simplex is actually an ideal simplex, then this map is no longer well-defined, because the edge between the two identified vertices gets sent to a vertex, which is not actually part of the 3-simplex. Assume, however, that that edge has been removed; then the map is again defined. Likewise, if we choose three vertices to be identified, then there is a unique linear degeneration of the 3-simplex into a line; on an ideal 3-simplex, the map is defined if the face determined by the three vertices is removed. See Figure 17 for details.

Our motivating philosophy is as follows. Thinking of Q as the union of two triangles with vertices $\pm\infty, v$ (v is on L), there is a unique linear degeneration of Q onto v sending $\pm\infty$ to v. This induces a homeomorphism $\dot{M} \to M$. Furthermore, it induces linear degenerations on all the tetrahedra from $\dot{\mathcal{S}}$ which intersect Q. Therefore some number of tetrahedra in $\dot{\mathcal{S}}$ degenerate, and we are left with a recipe with which to glue together the ones that remain. This should give a well-defined triangulation $\mathcal{S}$ of M because $\dot{\mathcal{S}}$ was constructed so that no tetrahedron was glued up to itself.

However, we need to be careful: as we have seen, not all linear degenerations of 3-simplices are well-defined for ideal tetrahedra. For example, if precisely two edges are removed (and no faces),

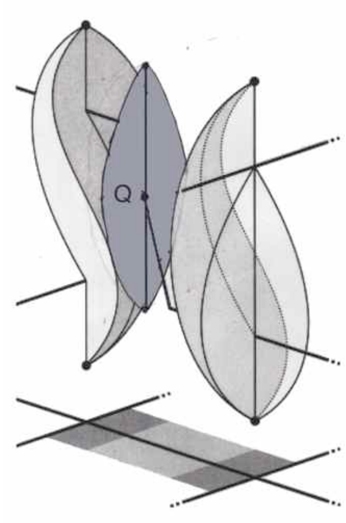

Figure 16:

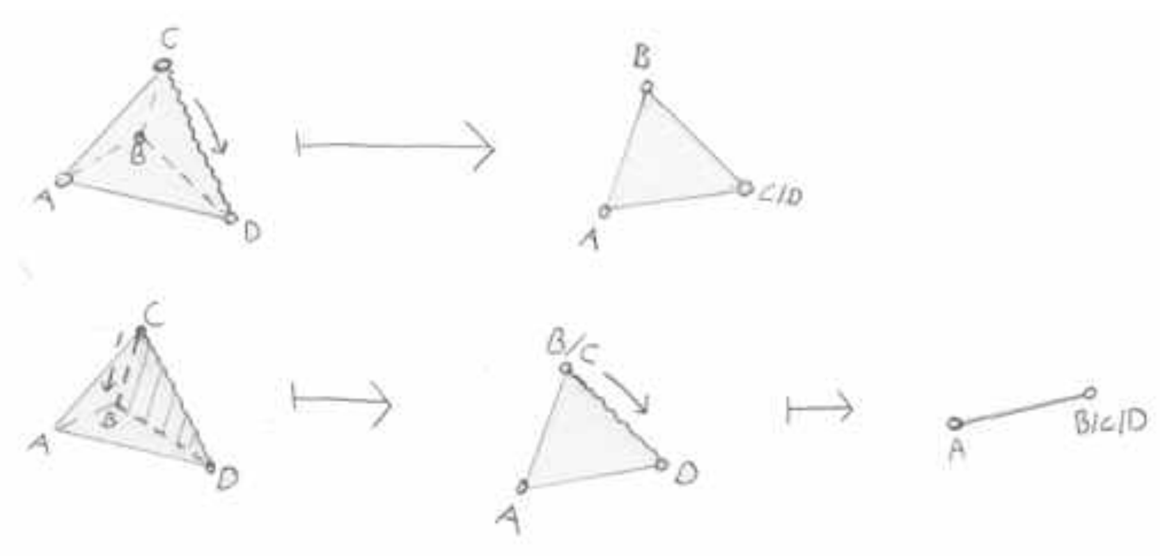

Figure 17:

then it is easy to see that we get a well-defined degeneration of an ideal simplex if and only if the two edges do not share a vertex.

It is not difficult to account for all the degenerations explicitly, from which one can check that they're all well-defined. We will describe the degenerations in terms of the octahedra O_ν. Note that the only octahedra intersecting Q in a face are O_1 and O_n. Furthermore, the edges contained in Q are $\dot{I}_{\varphi_-}$, $\dot{I}_{\varphi_+}$, $\dot{F}_0$, and $\dot{F}_{n+1}$. Now recall that $\mathcal{N}_\nu^*$ denotes the faces which abut vertex X_ν, which also correspond to the angles around X_ν, and therefore to the tetrahedra around $\dot{E}_\nu$. We will let $\mathcal{N}_\nu^\star$ denote the tetrahedra around X_ν which *survive* the degeneration.

The possibilities for intersecting Q are as follows:

1. O_ν such that ν is 1 or n. All four tetrahedra intersect Q. Two intersect in a face, so degenerate into a line; the other two intersect in an edge, either $\dot{I}_{\varphi_-}$ or $\dot{I}_{\varphi_+}$, and degenerate onto a single face. $\mathcal{N}_\nu^\star = \emptyset$.

2. O_ν such that $X_\nu \in \mathcal{N}_0, \mathcal{N}_1$ (i.e., X_ν belongs to the same pass as either X_1 or X_n). Then $Q \cap O_\nu$ is either $\dot{I}_{\varphi_+}$ or $\dot{I}_{\varphi_-}$. All four tetrahedra collapse onto a total of two faces. $\mathcal{N}_\nu^\star = \emptyset$.

3. O_ν such that $\nu = a, b$, *and* $\mathcal{N}_\nu$ contains R_0 or R_{n+1} (i.e., X_ν borders an unbounded face of the diagram). Then two tetrahedra intersect Q in the edge $\dot{I}_{\varphi_+}$ or $\dot{I}_{\varphi_-}$, and a third tetrahedron intersects Q in $\dot{F}_{n+1}$ or $\dot{F}_0$. $|\mathcal{N}_\nu^\star| = 1$.

4. O_ν such that $\nu = a, b$, but neither X_ν does not border an unbounded face of the diagram. Two tetrahedra intersect Q in the edge $\dot{I}_{\varphi_+}$ or $\dot{I}_{\varphi_-}$, and $|\mathcal{N}_\nu^\star| = 2$.

5. O_ν such that X_ν borders one of the unbounded faces, so that $O_\nu \cap Q$ consists either of $\dot{F}_0$ or $\dot{F}_{n+1}$. The tetrahedron corresponding to the unbounded angle is the only one that collapses, so $|\mathcal{N}_\nu^\star| = 3$.

6. $O_\nu \cap Q = \emptyset$, $|\mathcal{N}_\nu^\star| = 4$.

Recall that once we know how a tetrahedron intersects Q, we have a unique formula for its degeneration, simply by collapsing the intersection and applying the corresponding linear degeneration. In Figure 18, we've drawn the arc γ and the degeneration along it (i.e., the degeneration of the octahedra O_ν such that $X_\nu \in \mathcal{N}_1 \cup \{X_a, x_b\} \cup \mathcal{N}_0$.)

Here is a very nice way to describe the tetrahedra which have degenerated. Define a new diagram $G = D - \gamma$ (recall the γ is the arc joining X_a, p, and X_b) (Figure 19). G will be a 4-valent graph except at X_a and X_b, which will be 3-valent. Then the previous discussion implies that the tetrahedra which survive correspond exactly to the angles of G which do not border the unbounded region (where we do not count the "double angles" at the trivalent vertices).

7.4 Edges in $\mathcal{S}$

Now we have obtained an ideal topological triangulation $\mathcal{S}$ of M. The goal of this section is to write down the edges of $\mathcal{S}$ explicitly, so that we can then write down hyperbolic gluing equations for $\mathcal{S}$. In order to do this, we will use the diagram G introduce at the end of the previous section. Our notation for sets of edges and passes of G will be those we have already introduced, swapping G for D. (The over/under information at the trivalent vertices is still recorded, so that we can define the passes of G).

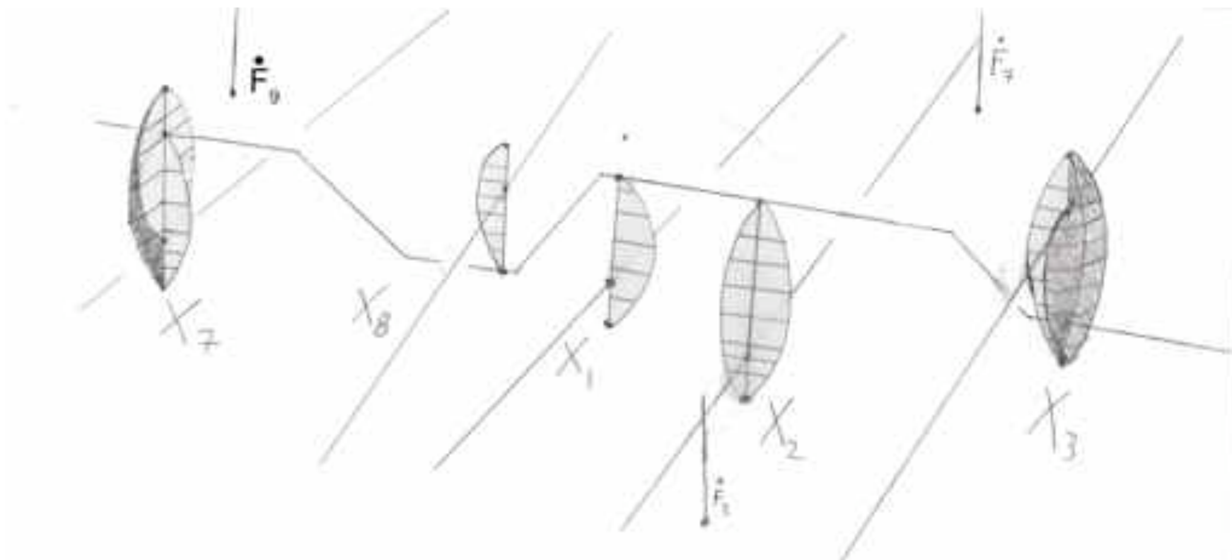

Figure 18: Crossings labeled as in Figure. 13. Remark that there are only 3 non-degenerate tetrahedra in the picture; two from $X_3 = X_a$, but only one from $X_7 = X_b$, because it is adjacent to $\dot{F}_9 = \dot{F}_{n+1}$, as we have drawn.

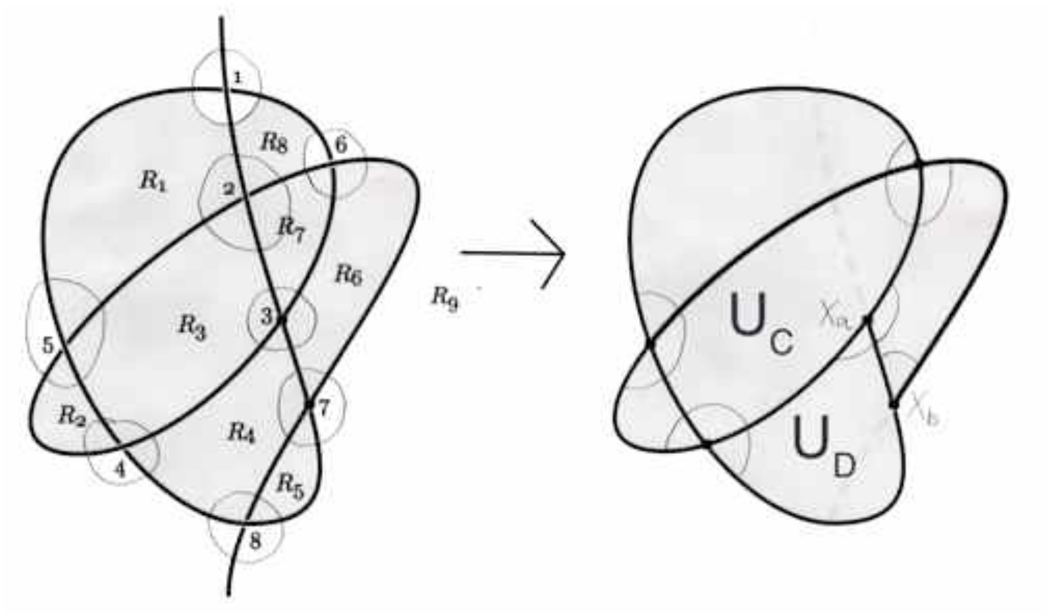

Figure 19: The angles indicated at the right correspond to the tetrahedra which have survived the degeneration. U_c and U_d are the faces which contain the two missing arcs, and abut the trivalent vertices.

Let $\{U_0, \ldots, U_t\}$ denote the faces of G; each will be the union of one or more faces of D. Without loss of generality, we assume U_0 is made up of the outer faces: $U_0 = R_0 \cup R_{n+1}$. Likewise, for each of the trivalent vertices X_a, X_b, there is a unique face of G which abuts that vertex and which contains the "double angle" (i.e., the face which contains the "missing" arc); let these faces be U_c and U_d, respectively (note: we may have $c = d$).

We recall that $\mathcal{E}_G$ denotes the set of edges in G (note that these are simply unions of edges from $\mathcal{E}_D$). For this section, we introduce a little more notation: suppose $\varphi \in \mathcal{P}_G$ is a pass of G. Then $\mathcal{E}_G^\varphi$ denotes the set of edges which intersect φ, and $\mathcal{E}_G^{\partial\varphi} \subset \mathcal{E}_G^\varphi$ denotes the two alternating edges which intersect φ at its beginning and end. $\mathcal{E}_G^0$ is the set of edges which intersects the outer face U_0. Finally, let $\alpha, \beta \in \mathcal{P}_G$ be the passes which "dead-end" at X_a and X_b, respectively (i.e., the passes which, if in D, would continue into the missing arc).

Theorem 7.3. The set of edges in $\mathcal{S}$ can be written as

$$\{E_\nu | X_\nu \notin \gamma\} \cup \{\, F_\tau | 1 \leq \tau \leq t, \tau \neq c, d\} \cup \{\, I_\varphi | \varphi \in \mathcal{P}_G\} \cup \bigcup_{\varphi \in \mathcal{P}_G} \left\{ H_\epsilon | \epsilon \in \mathcal{E}_G^\varphi - \left(\mathcal{E}_G^{\partial\varphi} \cup \mathcal{E}_G^0 \right) \right\} \tag{31}$$

where we set $I_\alpha = I_\beta$ if $c = d$. (Recall that $X_\nu \notin \gamma$ is the same as $X_\nu \notin \mathcal{N}_1 \cup \{a, b\} \cup \mathcal{N}_0$).

Theorem 7.3 is proved by explicitly examining the collapse of the tetrahedra in $\dot{\mathcal{S}}$. Here we give an overview. Beforehand, we make a remark about depictions of tetrahedra. Each of the octahedra O_ν, before collapse, has a *core*, by which we mean the set of four faces which intersect $\dot{E}_\nu$ (see Figure 21 for a good picture). In our figures, we will omit the outer faces, drawing only the core, so as to see the edges more clearly.

Now we begin. First, note that if $X_\nu \notin \gamma$, it is easy to see that $\dot{E}_\nu$ does not degenerate and gets sent to its own edge E_ν. Likewise, if $t \neq c, d$, then we just get the edges $\{\dot{F}_\mu | R_\mu \subset U_t\}$ degenerating to a single edge F_t (as has been illustrated in Figure 18; this is just the statement that the two edges $\dot{F}_7$ and $\dot{F}_3$ are getting identified "through" X_2 on γ).

Now, for an edge $\epsilon \in \mathcal{E}_G$, we define the following two sets of edges related to ϵ:

$$\mathcal{G}_\epsilon = \{\dot{E}_\nu | X_\nu \in \mathcal{N}_1, X_\nu \subset \epsilon\} \cup \{\, \dot{I}_\mu | \mu \in \mathcal{P}_D^+, (\mu \cap \epsilon) \neq \emptyset\} \cup \{\dot{H}_\eta | \eta \in \mathcal{K}_D^-, \eta \subset \epsilon\}$$

$$\bar{\mathcal{G}}_\epsilon = \{\dot{E}_\nu | X_\nu \in \mathcal{N}_0, X_\nu \subset \epsilon\} \cup \{\, \dot{I}_\mu | \mu \in \mathcal{P}_D^-, (\mu \cap \epsilon) \neq \emptyset\} \cup \{\dot{H}_\eta | \eta \in \mathcal{K}_D^+, \eta \subset \epsilon\}$$

Theorem 7.4. Let $\varphi \in \mathcal{P}_G^+$. Then the edges in $\bigcup_{\epsilon \in \mathcal{E}_G^\varphi} \mathcal{G}_\epsilon$ are sent to a single edge I_φ in $\mathcal{S}$. Further, the edges in $\bar{\mathcal{G}}_\epsilon$, for each $\epsilon \in \mathcal{E}_G^\varphi$, go to a single edge H_ϵ. If $\varphi \in \mathcal{P}_G^-$, then the edges in $\bigcup_{\epsilon \in \mathcal{E}_G^\varphi} \bar{\mathcal{G}}_\epsilon$ go to I_φ in $\mathcal{S}$, and the edges in $\mathcal{G}_\epsilon$, for each $\epsilon \in \mathcal{E}_G^\varphi$, go to H_ϵ.

Proof. First note that if $\varphi \in \mathcal{P}_G$ does not intersect γ, i.e., it is just a pass from D, then the theorem is obvious. To see this, note that in this case all edges ϵ such that $\epsilon \in \mathcal{E}_G^\varphi$ must also not intersect γ, i.e., they are just edges from D. Suppose that φ is an overpass. Then if $\epsilon \in \mathcal{E}_G^\varphi$ is non-alternating, we have $\mathcal{G}_\epsilon = \{\dot{I}_\varphi\}$, $\bar{\mathcal{G}}_\epsilon = \{\dot{H}_\epsilon\}$. If ϵ is alternating, then $\mathcal{G}_\epsilon = \{\dot{I}_\varphi\}$, $\bar{\mathcal{G}}_\epsilon = \{\dot{I}_\mu\}$ where μ is some underpass from D. So in this case, $\bigcup_{\epsilon \in \mathcal{E}_G^\varphi} \mathcal{G}_\epsilon$ contains only $\dot{I}_\varphi$, $\bar{\mathcal{G}}_\epsilon$ contains only $\dot{H}_\epsilon$ or $\dot{I}_\mu$, and the theorem just states that certain edges are identified to themselves (similarly if φ is an underpassing.)

If φ intersects γ, we must just write down the explicit edge degeneration. We give an example. In Figure 20, $\varphi \in \mathcal{P}_G^+$ is the overpassing of G which contains $\dot{E}_2$ (i.e. the overpassing which contains X_2). Let ϵ denote the edge of D between X_5 and X_2. Then we see that $\dot{H}_\epsilon$ collapses into $\dot{E}_2$ from one side (the dotted edge on the left of $\dot{E}_2$), and $\dot{I}_\varphi$ collapses to $\dot{E}_2$ from the right (the other dotted line) (just follow the arrows). Thus these edges are all identified in $\mathcal{S}$. Similar considerations for the other possible cases complete the proof.

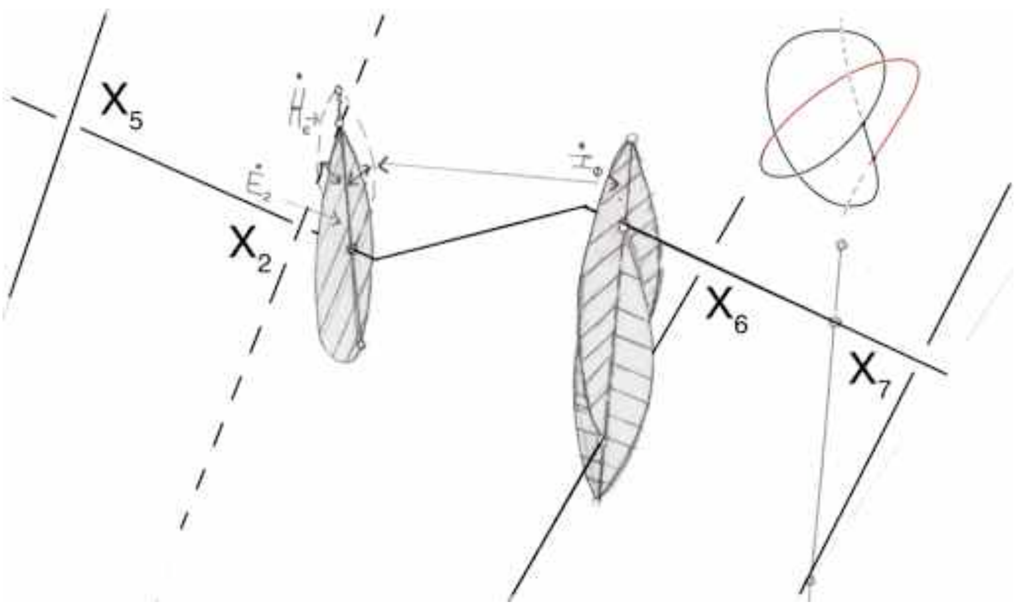

Figure 20: (note that the figure says ϕ instead of φ)

□

Note, however, (as we have seen explicitly in the case that $\epsilon \in \mathcal{E}_G$ is also in $\mathcal{E}_D$), that $H_\epsilon = I_\varphi$ for some $\varphi \in \mathcal{P}_G$ if ϵ is an alternating edge, i.e., $\epsilon \in \mathcal{E}_G^{\partial\varphi}$.

Now we account for the collapse of the edges $\dot{F}_0, \dot{F}_{n+1}$.

Theorem 7.5. If $\epsilon \in \mathcal{E}_G$ is a non-alternating edge such that $\epsilon \in \mathcal{E}_0$, then the edge H_ϵ given in Theorem 7.4 is actually equal to I_φ, where I_φ, where φ pass containing ϵ.

Proof. As before, we see how our triangulation degenerates in response to the degeneration of $\dot{F}_0$ and $\dot{F}_{n+1}$. In particular, the claim is simply the statement that, when the red curve ($\dot{F}_0$ or $\dot{F}_{n+1}$) is collapsed in Figure 21, the identification on the edges induced by the linear degeneration of the tetrahedra sends $\dot{H}_g \to \dot{I}_g$, and $\dot{I}_f \to \dot{H}_f$.

□

Now remark that
$$\{\dot{E}_\nu | \nu \notin \gamma\} \cup \{\dot{F}_\mu | R_\mu \not\subset U_c, U_d\} \cup \bigcup_{\epsilon \in \mathcal{E}_G} (\mathcal{G}_\epsilon \cup \bar{\mathcal{G}}_\epsilon)$$

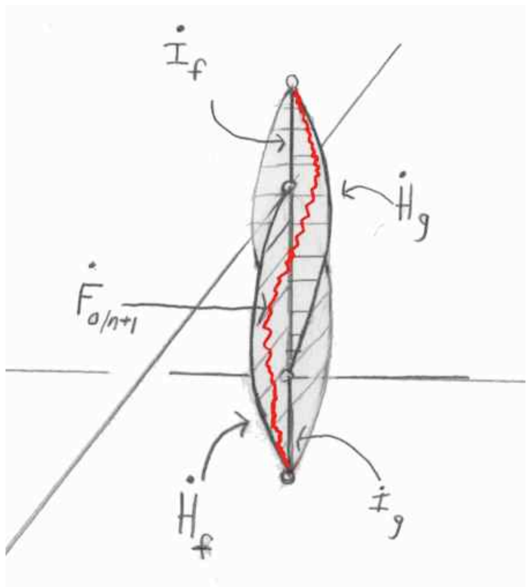

Figure 21:

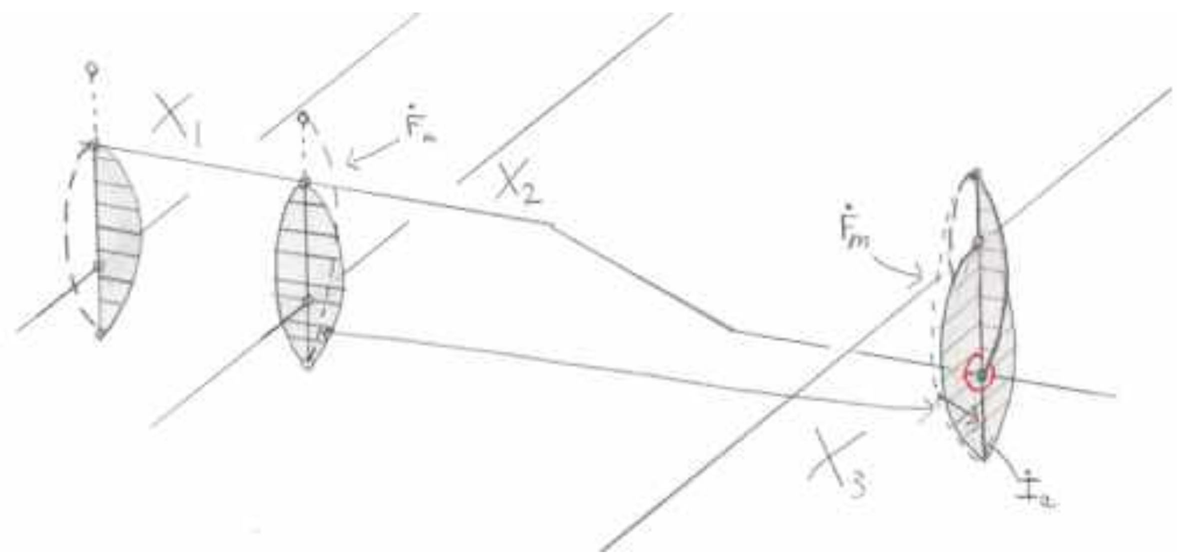

Figure 22: $\dot{F}_m$, in U_c, collapsing to I_α.

consists of all the edges of $\dot{\mathcal{S}}$ other than $\dot{E}_a, \dot{E}_b$, and $\{\dot{F}_\mu | \mu \subset U_c \cup U_d\}$. These go either to I_α or I_β, as shown in Figure 22.

Thus we have accounted for all the edges in $\dot{\mathcal{S}}$, and one can go back and check that we have gotten precisely the edges from Theorem 7.3:

$$\{E_\nu | \nu \notin \mathcal{N}_1 \cup \{a,b\} \cup \mathcal{N}_0\} \cup \{F_\tau | 1 \leq \tau \leq t, t \neq c,d\} \cup \{I_\varphi | \varphi \in \mathcal{P}_G\} \cup \Big(\cup_{\varphi \in \mathcal{P}_G} \{H_\epsilon | \epsilon \in \mathcal{E}_G^\varphi - \{\mathcal{E}_G^{\partial\varphi} \cup \mathcal{E}_G^0\}\} \Big)$$

Lemma 7.3. *None of these edges are homotopically trivial.*

Recall from §6 that this means that when we lift each edge via π^{-1}, ($\pi : \mathbb{H}^3 \to M$ being a covering map determining the complete hyperbolic structure on M), none of the edges lift to a loop in $\bar{\mathbb{H}}^3$. It turns out that this is equivalent to saying that none of the edges of $\mathcal{S}$ are homotopic to a portion of the knot. Yokota gives an incomplete proof that this is the case. The author intends to go through it fully in the near feature. Assuming Lemma 7.3 is true, then, by the results of §6, the hyperbolic gluing equations for $\mathcal{S}$ have a unique solution. We shall assume this in what follows.

7.5 Hyperbolic Gluing Equations

In this section we write down the hyperbolic gluing equations for $\mathcal{S}$ and define the potential function $V(z)$ whose imaginary part at a critical point gives the hyperbolic volume of M.

For convenience, we recall the edges of $\mathcal{S}$:

$$\{E_\nu | X_\nu \notin \gamma\} \cup \{\, F_\tau | 1 \leq \tau \leq t, \tau \neq c,d\} \cup \{\, I_\varphi | \varphi \in \mathcal{P}_G\} \cup \bigcup_{\varphi \in \mathcal{P}_G} \Big\{ H_\epsilon | \epsilon \in \mathcal{E}_G^\varphi - \Big(\mathcal{E}_G^{\partial\varphi} \cup \mathcal{E}_G^0 \Big) \Big\}$$

with $I_\alpha = I_\beta$ if $c = d$.

The tetrahedra in $\mathcal{S}$ are those from $\dot{\mathcal{S}}$ which survive the degeneration, and therefore we can use the same notation $S_{\nu\mu}$ to refer to them (recall that $S_{\nu\mu}$ projects onto the angle (ν, μ) of D). Recall that, to "straighten out" a non-degenerate ideal topological triangulation of a complete hyperbolic 3-manifold, we assign each tetrahedron a modulus, corresponding to a choice of edge (opposite edges giving the same moduli), and then we can obtain a set of equations that they must satisfy. Each tetrahedron $S_{\nu\mu}$ in $\mathcal{S}$ has a pair of opposite edges E_ν, F_μ. Let $z_{\nu\mu}$ denote the modulus corresponding to these edges. As we have seen, the tetrahedra in $\mathcal{S}$ correspond precisely to the angles of G (minus those angles which either border the unbounded face U_0, or those that form a "double" angle around one of the trivalent vertices of G). Therefore, letting $z : \mathcal{E}_G \to \mathbb{C}$ be some function, we can use this function to assign a modulus to each tetrahedron in $\mathcal{S}$ by letting $z_{\nu\mu} = z(\epsilon)/z(\eta)$, see Figure 23:

A solution to $\mathcal{H}$ obtained in this manner will be called *edge determined*.

Theorem 7.6. *Any solution to $\mathcal{H}$ must be edge determined, by a function $z : \mathcal{E}_G \to \mathbb{C}$ such that $z(\mathcal{E}_G^0) = 1$.*

Proof. To prove the theorem, we show there is a subset of equations in $\mathcal{H}$ which are satisfied if and only if the moduli are edge-determined by such a function z.

In particular, we will consider the edge equations corresponding to $\{E_\nu | \nu \notin \gamma\}, \{F_\tau | \tau \neq c,d\}, I_\alpha, I_\beta$, and two cusp equations. Recall that $\mathcal{N}_\nu^\star$ denotes the faces R_μ corresponding to the tetrahedra around X_ν that did not degenerate, that $\mathcal{R}_\mu$ is the set of crossings bordering R_μ, and

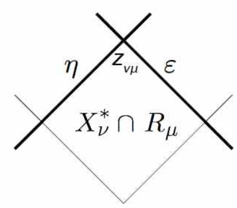

Figure 23:

that $\mathcal{N}_1$ and $\mathcal{N}_0$ are the vertices in $\gamma, \neq X_a, X_b$. The edge equations around E_ν are easy to write down:

$$\prod_{R_\mu \in \mathcal{N}_\nu^\star} z_{\nu\mu} = 1$$

Clearly all solutions to these equations are edge-determined: just go in a circle around the given crossing, and set the value of z at each edge to be the ratio of adjacent moduli; that the value of this function z so obtained can be chosen to be 1 on $\mathcal{E}_G^0$ is due exactly to the fact that there is no modulus at an unbounded angle, because the tetrahedron there has degenerated. Likewise, around F_τ, $\tau \neq c, d$, we get the equation

$$\prod_{\mu | R_\mu \subset U_\tau} \left\{ \prod_{\nu \in \mathcal{R}_\mu - (\mathcal{N}_1 \cup \mathcal{N}_0)} z_{\nu\mu} \right\} = 1,$$

and it is equally easy to see these are edge determined—now we go in a circle around the boundary of the given face U_t, again setting the value of z on an edge bordering U_t to be the ratio of adjacent moduli. However, if we do this around U_c or U_d, we will run into a problem at the trivalent vertex, which does not correspond to any moduli in the interior or $U_{c/d}$. Indeed, suppose that z_1, z_2 are both moduli around X_a which have not degenerated (the angles they correspond to are outside of U_c). Then, if we try to edge-determine the moduli, the product around U_c will be $z_1 z_2$. Thus the following Lemma proves Theorem 7.6:

Lemma 7.4. We can replace the edge equations around I_α and I_β by

$$\prod_{\mu | R_\mu \subset U_c} \left\{ \prod_{\nu \in \mathcal{R}_\mu - (\mathcal{N}_1 \cup \mathcal{N}_0)} z_{\nu\mu} \right\} = \prod_{\mu \in \mathcal{N}_a^\star} z_{a\mu}, \quad \prod_{\mu | R_\mu \subset U_d} \left\{ \prod_{\nu \in \mathcal{R}_\mu - (\mathcal{N}_1 \cup \mathcal{N}_0)} z_{\nu\mu} \right\} = \prod_{\mu \in \mathcal{N}_b^\star} z_{b\mu}$$

if $c \neq d$ and with

$$\prod_{\mu | R_\mu \subset U_c} \left\{ \prod_{\nu \in \mathcal{R}_\mu - (\mathcal{N}_1 \cup \mathcal{N}_0)} z_{\nu\mu} \right\} \cdot \prod_{\mu | R_\mu \subset U_d} \left\{ \prod_{\nu \in \mathcal{R}_\mu - (\mathcal{N}_1 \cup \mathcal{N}_0)} z_{\nu\mu} \right\} = \prod_{\mu \in \mathcal{N}_a^\star} z_{a\mu} \cdot \prod_{\mu \in \mathcal{N}_b^\star} z_{b\mu}$$

if $c = d$.

Proof. First recall that the edges $\dot{F}_\mu$ have collapsed to I_α, when $R_\mu \subset U_c$ (and likewise for I_β and $R_\mu \subset U_d$.) Thus the edge equations around I_α, I_β can be written as

$$x \cdot \prod_{\mu | R_\mu \subset U_c} \left(\prod_{\nu \in \mathcal{R}_\mu - (\mathcal{N}_1 \cup \mathcal{N}_2)} z_{\nu\mu} \right) = 1, \quad y \cdot \prod_{\mu | R_\mu \subset U_d} \left(\prod_{\nu \in \mathcal{R}_\mu - (\mathcal{N}_1 \cup \mathcal{N}_2)} z_{\nu\mu} \right) = 1 \tag{32}$$

if $c \neq d$ and as

$$xy \cdot \prod_{\mu | R_\mu \subset U_c} \left(\prod_{\nu \in \mathcal{R}_\mu - (\mathcal{N}_1 \cup \mathcal{N}_2)} z_{\nu\mu} \right) \cdot \prod_{\mu | R_\mu \subset U_d} \left(\prod_{\nu \in \mathcal{R}_\mu - (\mathcal{N}_1 \cup \mathcal{N}_2)} z_{\nu\mu} \right) = 1 \tag{33}$$

if $c = d$, where x and y are complex numbers corresponding to the other moduli around I_α and I_β, respectively.

To rewrite these equations, we will combine them with certain cusp equations. In particular, recall that the two tetrahedra at X_a which were facing X_1 degenerated onto a bigon, see Figure 22. In what follows, let $\dot{\mathcal{L}}$ denote a regular neighborhood of L triangulated by $\dot{\mathcal{S}}$, and $\mathcal{L}$ the same neighborhood triangulated by $\mathcal{S}$. The red loop drawn in Figure 22, and blown up in Figure 24, is a simplicial loop on $\mathcal{L}$, (we have also drawn it in red on the picture of $\dot{\mathcal{L}}$ in Figure 26). Oriented as indicated, the moduli at its right are those that correspond to the edge E_ν of the tetrahedra in $\mathcal{N}_a^\star$, as well as any moduli belonging to the edges which get identified to I_α from the right. The only edges identified to I_α from the left are $\dot{F}_\mu$ with $R_\mu \subset U_c$, so it follows that the equation from this cusp is

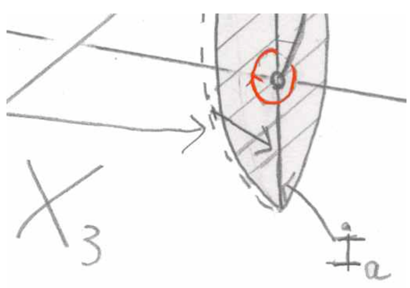

Figure 24:

$$x \cdot \prod_{\mu \in \mathcal{N}_a^\star} z_{a\mu} = 1$$

with x as in (32). Likewise, for X_b and U_d, we get

$$y \cdot \prod_{\mu \in \mathcal{N}_b^\star} z_{b\mu} = 1$$

with y as in (32). These substitute into (32) to give the desired result.

□

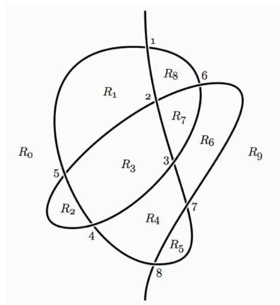

Figure 25:

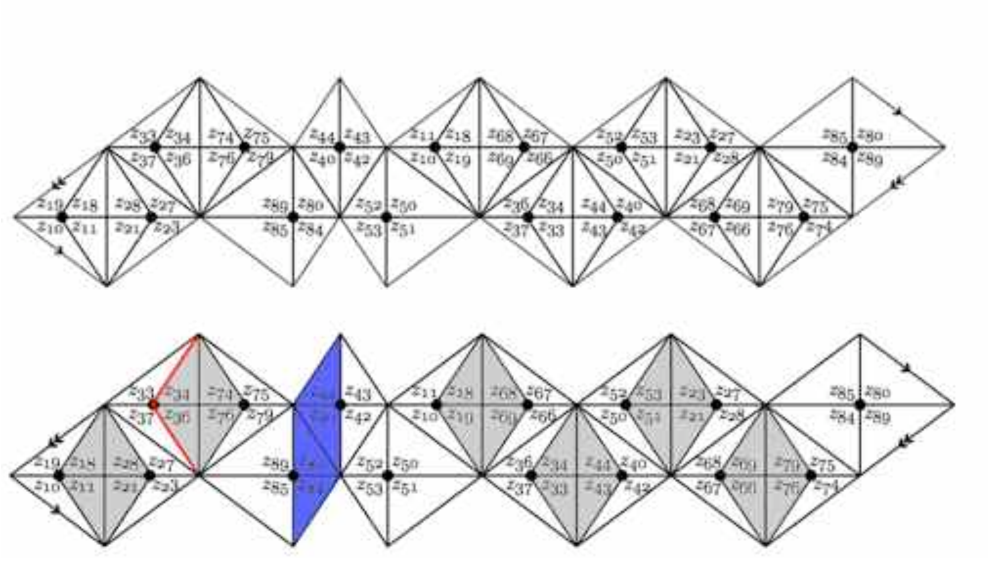

Figure 26: Triangulation of the knot L by $\dot{\mathcal{S}}$

□

It is evident how we obtained the triangulation $\dot{\mathcal{L}}$ in Figure 26. Remark that to go from $\dot{\mathcal{L}}$ to $\mathcal{L}$, we need to glue the triangulations of regular neighborhoods of $\pm\infty$ to $\dot{\mathcal{L}}$ along Q; this can be described explicitly, see [29]. We then collapse the edges in $\dot{\mathcal{L}}$ which correspond to faces of tetrahedra that have degenerated. Note, however, that for the edges H_ϵ, the entire edge equations can be read from $\dot{\mathcal{L}}$. Remark that any edges of $\dot{\mathcal{S}}$ intersecting $\dot{\mathcal{L}}$ of correspond to vertices on $\dot{\mathcal{L}}$, surrounded by some number of triangles corresponding to tetrahedra. We call this situation the *star* of the edge. Now, by Theorem 7.6, without loss of generality, we can choose a function $z : \mathcal{E}_G \to \mathbb{C}$ such that $z(\mathcal{E}_G^0) = 1$, and assume that it determines the moduli of our triangulation. We will read some more edge equations of $\mathcal{S}$, from $\dot{\mathcal{L}}$, and deduce a potential function for these equations.

We write $\operatorname{sgn}(\nu,\mu) = 1$ if ϵ crosses over η, and $\operatorname{sgn}(\nu,\mu) = -1$ if ϵ crosses under η, in Figure 23. Given an angle (ν,μ), write $\epsilon_{\nu\mu}$ and $\eta_{\nu\mu}$ for the corresponding edges.

Theorem 7.7. Let

$$V(z) = \sum_{\nu\in\mathcal{N}}\sum_{\mu\in\mathcal{N}_\nu^\star} sgn(\nu,\mu) Li_2\left(z_{\nu\mu}^{sgn(\nu,\mu)}\right) = \sum_{\nu\in\mathcal{N}}\sum_{\mu\in\mathcal{N}_\nu^\star} sgn(\nu,\mu) Li_2\left(\left(\frac{z(\epsilon_{\nu\mu})}{z(\eta_{\nu\mu})}\right)^{sgn(\nu,\mu)}\right)$$

Then when z_0 is a labeling which determines a solution to the hyperbolic edge equations, we have $\frac{\partial V_0(z)}{\partial z(\epsilon)}|_{z_0} = 0$ for some branch of V. Furthermore, $Im(V(z_0)) = Vol(M)$.

Proof. Suppose we have an edge $\epsilon \in \mathcal{E}_G - \mathcal{E}_G^0$, surrounded by edges e, f, g, h as in Figure 27. In Figure 26, we have colored in gray the stars corresponding to H_ϵ, ϵ non-alternating. As mentioned, these will descend to stars of $\mathcal{L}$, possibly with some edges collapsed.

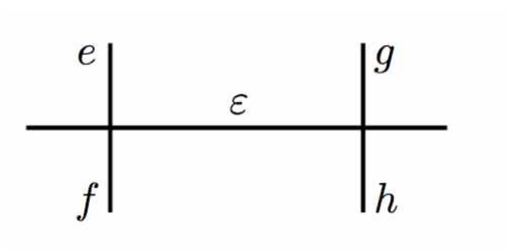

Figure 27:

If ϵ is contained in an overpass, and $X_a \notin \epsilon$, then the star of H_ϵ in $\mathcal{L}$ looks like 28, and the edge relations around H_ϵ can be read as

$$\frac{1 - z(f)/z(\epsilon)}{1 - z(e)/z(\epsilon)} \cdot \frac{1 - z(g)/z(\epsilon)}{1 - z(h)/z(\epsilon)} = 1$$

Letting

$$V_\epsilon = \mathrm{Li}_2(z(e)/z(\epsilon)) - \mathrm{Li}_2(z(f)/z(\epsilon)) - \mathrm{Li}_2(z(g)/z(\epsilon)) + \mathrm{Li}_2(z(h)/z(\epsilon)) \pm 2\pi i \log z(\epsilon)$$

we see that $\partial V_\epsilon(z)/\partial z(\epsilon)$ vanishes at the hyperbolic edge solution, because the dihedral angles around the edge H_ϵ must add up to 2π (the author is unsure at present about the details of the

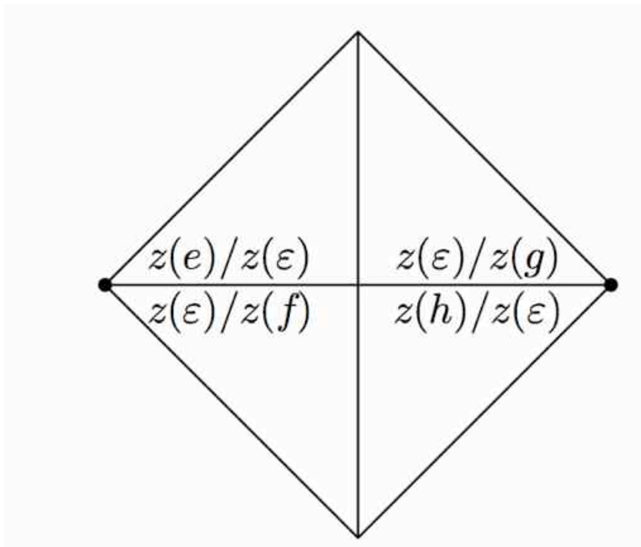

Figure 28:

sign choice, but it should not be difficult to uncover). Otherwise, one simply takes the derivative, using the defintion of Li_2 as an integral from §6, to verify the above claim.

If $X_a \in \epsilon$, then one of the edges η from $\{e, f, g, h\}$ does not exist, as there are only 3 tetrahedra abutting H_ϵ. Therefore we simply set $z(\eta) = 0$ above.

The situation is similar when ϵ is contained in an underpass. If $X_b \notin \epsilon$, then the star of H_ϵ looks like Figure 29, and the equation we get is

$$\frac{1 - z(\epsilon)/z(e)}{1 - z(\epsilon)/z(f)} \cdot \frac{1 - z(\epsilon)/z(h)}{1 - z(\epsilon)/z(g)} = 1$$

whose potential function is

$$V_\epsilon = -\mathrm{Li}_2(z(\epsilon)/z(e)) + \mathrm{Li}_2(z(\epsilon)/z(f)) + \mathrm{Li}_2(z(\epsilon)/z(g)) + \mathrm{Li}_2(z(\epsilon)/z(h)) \pm 2\pi i \log z(\epsilon)$$

Again, if $X_b \in \epsilon$, an edge η from $\{f, e, g, h\}$ does not exist, and we set $z(\eta) = \infty$.

Now suppose ϵ is an alternating edge, between the overpass I_φ and the underpass I_γ. A piece of the star of these two edges in $\mathcal{L}$ is shown in Figure 30, and highlighted in blue in 26. We have also drawn two meridians, in red. Note that $z_1 z_2 = -1/z_3$ for the three moduli of a tetrahedron. Then we write down two cusp equations, one per each of the red meridians, along the inside of the meridian (the side facing the other one). It is easy to reduce this to the equation

$$\frac{1 - z(f)/z(\epsilon)}{1 - z(e)/z(\epsilon)} = \frac{1 - z(\epsilon)/z(h)}{1 - z(\epsilon)/z(g)}$$

whose potential function is

$$V_\epsilon = \mathrm{Li}_2(z(e)/z(\epsilon)) - \mathrm{Li}_2(z(f)/z(\epsilon)) + \mathrm{Li}_2(z(\epsilon)/z(g)) - \mathrm{Li}_2(z(\epsilon)/z(h))$$

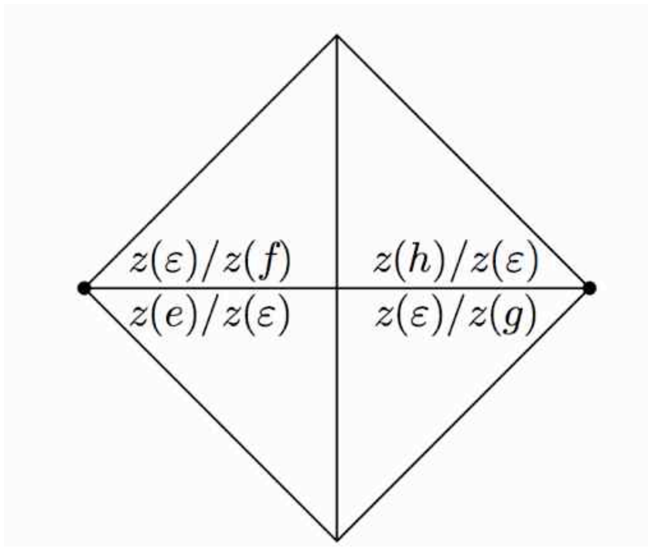

Figure 29:

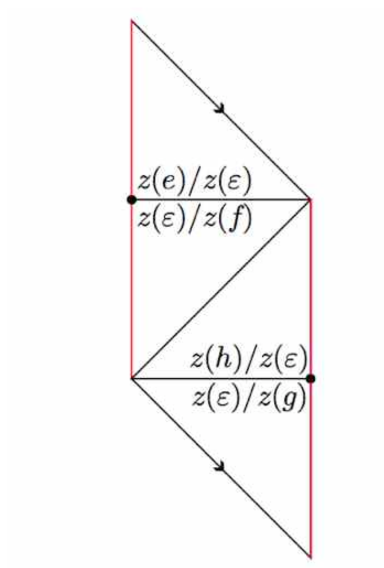

Figure 30:

Now, summing all the distinct terms from all the V_ϵ, we get

$$V(z) := \sum_{\nu \in \mathcal{N}} \sum_{\mu \in \mathcal{N}_\nu^\star} \operatorname{sgn}(\nu, \mu) \mathrm{Li}_2 \left(\left(\frac{z(\epsilon_{\nu\mu})}{z(\eta_{\nu\mu})} \right)^{\operatorname{sgn}(\nu,\mu)} \right)$$

Therefore, a branch of V has a critical point at the solution to the hyperbolic gluing equations (note that changing the branch of a dilogarithm $\mathrm{Li}_2(z)$ corresponds to adding a term of the form $\pm 2\pi i \log(z)$). Finally, it is easy to compute

$$\mathrm{Im}V(z) = \sum_{\nu \in \mathcal{N}} \sum_{\mu \in \mathcal{N}^\star} D(z_{\nu\mu}) + \sum_{\epsilon \in \mathcal{E}_G - \mathcal{E}_{G0}} \log|z(\epsilon)| \cdot \mathrm{Im}z(\epsilon) \frac{\partial V(z)}{\partial z(\epsilon)}$$

where $D(z) = \mathrm{ImLi}_2(z) + \log|z| \arg(1-z)$ is the Bloch-Wigner function from §6. Thus, at a critical point, we have $\mathrm{Im}V(z) = \sum_{\nu \in \mathcal{N}} \sum_{\mu \in \mathcal{N}^\star} D(z_{\nu\mu}) = \mathrm{Vol}(M)$, as desired. $\square$

7.6 Computing the Colored Jones Polynomial

Now that we have defined the potential function $V(z)$, we just need to compute $F'_{N-1}(L)$ and show that is has the form in (30).

We recall the state-sum definition of $F'_{N-1}(L)$, which we will use to compute it. We take the graph D, orient it, and insert a set $\mathcal{X}$ of vertices at its maxima and minima, with signs as in Figure 31. This new graph, with more vertices and edges, will still be denoted D. We let $\mathcal{N}$ denote the set of vertices of D which come from crossings, and $\dot{\Psi}$ the set of functions (states) $\sigma : \mathcal{E}_D \to \{0, \ldots, N-1\}$, which take the value 0 on the broken edge, from X_1 to X_n, intersecting $p \in D$.

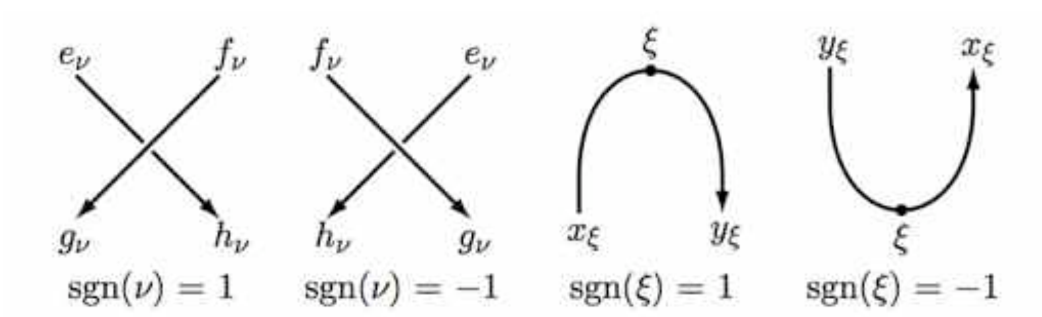

Figure 31: Signs for vertices

Then we have

$$F_{N-1}(L') = \prod_{\xi \in \mathcal{X}} -q^{\operatorname{sgn}(\xi)/2} \cdot \sum_{\sigma \in \dot{\Psi}} \left(\prod_{\nu \in \mathcal{N}} \langle D|\sigma \rangle_\nu \prod_{\xi \in \mathcal{X}} \delta_{\sigma(x_\xi)+1, \sigma(y_\xi)} \right)$$

where

$$\langle D|\sigma\rangle_\nu = \begin{cases} (S)^{\sigma(e_\nu)\sigma(f_\nu)}_{\sigma(g_\nu)\sigma(h_\nu)} & \text{if } \mathrm{sgn}(\nu)=1 \\ (\bar{S})^{\sigma(f_\nu)\sigma(e_\nu)}_{\sigma(h_\nu)\sigma(g_\nu)} & \text{if } \mathrm{sgn}(\nu)=-1 \end{cases}$$

with

$$\begin{aligned}
(S)^{ij}_{kl} &= Nq^{-1/2-(k-j)(i-l+1)} \frac{\theta^{ij}_{kl}}{(\bar{q})_{i-j}(q)_{j-l}(\bar{q})_{l-k-1}(\bar{q})_{k-i}} \\
(\bar{S})^{ij}_{kl} &= Nq^{1/2+(i-l)(k-j+1)} \frac{\theta^{ij}_{kl}}{(q)_{i-j}(\bar{q})_{j-l}(q)_{l-k-1}(\bar{q})_{k-i}} \\
[\mathrm{m}] &= \text{residue of m in } \{1,\dots N\} \\
(w)_m &= (1-w)(1-w^2)\dots(1-w^{[m]}), \text{ called the } \textit{quantum factorial}. \\
\theta^{ij}_{kl} &= \begin{cases} 1 & \text{if } [i-j]+[j-l]+[l-k-1]+[k-i]=N-1, \\ 0 & \text{otherwise.} \end{cases} \\
q &= e^{2\pi i/N}
\end{aligned}$$

and the overbar denoting inversion. Furthermore, by the *wedge* of $X_\nu \in \mathcal{N}$ we will mean the face R_μ of D which intersects the angle (ν,μ) of X_ν between the edges g_ν and h_ν (see Figure 31).

Remark 7.2. Again, note that $\theta = 1$ if and only if q^i, q^j, q^l, q^k go around the unit circle clockwise, and $l \neq k$.

For each $\sigma \in \dot{\Psi}$ and choice (ν,μ) of angle of D, define (with ϵ, η as in Figure 23):

$$\sigma(\nu,\mu) = \begin{cases} [\sigma(\epsilon)-\sigma(\eta)-1] \text{ if } R_\mu \text{ is the wedge of } X_\nu \\ [\sigma(\epsilon)-\sigma(\eta)] \text{ otherwise} \end{cases}$$

Then, using the identity

$$(q)_s = \pm(-1)^s q^{s(s+1)/2}(\bar{q})_s,$$

we can rewrite $\langle D|\sigma\rangle_\nu$ as

$$\langle D|\sigma\rangle_\nu = \pm N q^{\sigma(e_\nu - g_\nu)-\mathrm{sgn}(\nu)/2} \prod_{\mu\in\mathcal{N}^*_\nu} \frac{1}{\left(q^{\mathrm{sgn}(\nu,\mu)}\right)_{\mathrm{sgn}(\nu,\mu)\sigma(\nu,\mu)}} \tag{34}$$

up to a factor of θ. Thus, *a priori*, to compute $F'_{N-1}(L)$ we are taking the sum over indices of a large product which has one quantum factorial for each angle of L, i.e., one factorial for each tetrahedron in $\dot{\mathcal{S}}$. Now, observe that, for large N,

$$\begin{aligned}
\frac{1}{(q)_{[s]}} = \exp\left(-\frac{N}{2\pi}\sum_{k=1}^{s}\frac{2\pi}{N}\log(1-q^k)\right) &\approx \chi(s,N)\cdot\exp\left(-\frac{N}{2\pi}\int_0^{2\pi s/N}\log\left(1-e^{xi}\right)dx\right) \\
&= \chi(s,N)\cdot\exp\left(-\frac{N}{2\pi}\int_1^{q^s}\frac{\log(1-y)}{y}dy\right) \\
&= \chi(s,N)\cdot\exp\left(\mathrm{Li}_2(q^s)-\frac{\pi^2}{6}\right) \\
&= \chi(s,N)\cdot\exp\left(\mathrm{Li}_2(q^s)\right)
\end{aligned}$$

(we have absorbed the polynomial-growth terms into χ). Thus each quantum factorial seems to be contributing the exponential of a dilogarithm to $F'_{N-1}(L)$. This is exactly what we need to get $e^{V(z)}$, except that we only want dilogarithms for the tetrahedra which survive the degeneration from $\dot{\mathcal{S}}$ to $\mathcal{S}$. Note, however, that certainly not all states contribute to the state-sum; a state σ *contributes* if and only if

$$\langle D|\sigma\rangle := \prod_{\nu\in\mathcal{N}} \langle D|\sigma\rangle_\nu \prod_{\xi\in\mathcal{X}} \delta_{\sigma(x_\xi)+1,\sigma(y_\xi)} \neq 0$$

Therefore we might hope, after accounting for the non-contributing states, to eliminate from $F'_{N-1}(L)$ the factorials associated to degenerating tetrahedra, and indeed, somewhat miraculously, this is possible. Thus the dilogarithms we are left with correspond precisely to the tetrahedra in $\mathcal{S}$.

First we have

Lemma 7.5. *Suppose σ is contributing. Then*

1. *$\sigma(\nu,\mu) = 0$ if the angle (ν,μ) is unbounded, i.e. $\mu \in \{0, n+1\}$.*
2. *$\sum_{\nu\in\mathcal{R}_\mu} = N-1$ if R_μ is bounded, i.e., $\mu \notin \{0, n+1\}$ (the sum of angles around a bounded face is $N-1$).*
3. *$\sum_{\mu\in Q_\nu} \sigma(\nu,\mu) = N-1$ (the sum of angles around a crossing is $N-1$).*

Proof. First, we prove

Lemma 7.6. *$\sum_{\nu\in\mathcal{R}_\mu} \sigma(\nu,\mu) \equiv -1 \mod N$ unless $\mu \in \{0, n+1\}$.*

Proof. Give ∂R_μ the counter-clockwise orientation, and let v_ξ, for $\xi \in \mathcal{X} \cap \partial R_\mu$, denote the vector tangent to ∂R_μ at ξ. Then, according as the orientation of D at ξ agrees with the counter clockwise orientation of R_μ, we have $v_\xi = u_\xi$ or $v_\xi = -u_\xi$. Call ξ *positive* in the former case, and *negative* in the latter; then ξ is called a positive or negative maximum or minimum according as ξ is a local maximum or minimum. Let $p_\mu, \overline{p}_\mu, q_\mu, \overline{q}_\mu$ denote the number of positive maxima, negative maxima, positive minima, and negative minima of R_μ, respectively. Furthermore, let r_μ denote the number of $\nu \in \mathcal{R}_\mu$ such that R_μ is the wedge of X_ν. Then as we go around the crossings ν in ∂R_μ, summing $\sigma(\nu,\mu)$, we will get a telescoping sum, except for the fact that there are some wedges, and that at each ξ, the label value must change by one (because of the $\delta_{\sigma(x_\xi)+1,\sigma(y_\xi)}$ terms, and that we are assuming the state is contributing). Therefore, it can be seen directly that

$$\sum_{\nu\in\mathcal{R}_\mu} \equiv -p_\mu + \overline{p}_\mu + q_\mu - \overline{q}_\mu - r_\mu \mod N$$

where the right-hand side counts the number of times ∂R_μ goes around R^*_μ in $\mathbb{R}^2$ in the counter-clockwise direction, so its -1. ☐

To prove Lemma 7.5, first note that (3) follows immediately from the definition of θ and the assumption that σ is contributing. Furthermore, $\sum_{\nu=1}^{n} \sum_{\mu\in Q_\nu} \sigma(\nu,\mu) = \sum_{\mu=0}^{n+1} \sum_{\nu\in\mathcal{R}_\mu} \sigma(\nu,\mu)$. Therefore

$$\sum_{\mu=0}^{n+1} \sum_{\nu\in\mathcal{R}_\mu} \sigma(\nu,\mu) = n(N-1)$$

But $\sigma(\nu,\mu) \geq 0$, so by Lemma 7.6, $\sum_{\nu\in\mathcal{R}_\mu} \sigma(\nu,\mu) \geq N-1$ when $\mu \notin \{0, n+1\}$, giving us parts (1) and (2) of Lemma 7.5. $\square$

Thus we already see that we have eliminated the contributions from the tetrahedra bordering the unbounded faces R_0 and R_{n+1}; recall that these tetrahedra degenerate because they contain edges $\dot{F}_{n+1}$ or $\dot{F}_1$.

The technical core of the rest of the proof relies on some q-polynomial identities, which are true when q is an Nth root of unity, as we are assuming. We write $i \in [j,k]$ to indicate that q^j, q^i, and q^k go clockwise around the unit circle, i.e. $[i-k]+[k-j]=[i-j]$.

We have

Lemma 7.7.

$$\sum_{i\in[k,j]} q^{-i}\bar{S}^{ij}_{kl} = \delta_{j,k}q^{1-l}, \ \sum_{j\in[i,l]} q^{-j}S^{ij}_{kl} = \delta_{i,l}q^{-1-k}, \ \sum_{k\in[l,i]} q^{k}\bar{S}^{ij}_{kl} = \delta_{i+1,l}q^{j}, \ \sum_{l\in[j,k]} q^{l}S^{ij}_{kl} = \delta_{j,k+1}q^{i}$$

and

Lemma 7.8.

$$\sum_{i\in[k,j]} q^{-i}S^{ij}_{kl} = \frac{-Nq^{-1-k}}{(\bar{q})_{[j-l]}(q)_{[l-k-1]}}, \ \sum_{j\in[i,l]} q^{-j}\bar{S}^{ij}_{kl} = \frac{-Nq^{1-l}}{(\bar{q})_{[l-k-1]}(q)_{[k-i]}},$$

$$\sum_{k\in[j,l]} q^{k}S^{ij}_{kl} = \frac{-Nq^{-1+i}}{(q)_{[i-j]}(\bar{q})_{[j-l]}}, \ \sum_{l\in[k,i]} q^{l}\bar{S}^{ij}_{kl} = \frac{-Nq^{1+j}}{(\bar{q})_{[i-j]}(q)_{[k-i]}}.$$

The proofs can be found in [29] and [17].

Recall that γ denotes the "broken arc" of D, connecting X_a and X_b through p (we may assume γ contains no maxima or minima), and that $\mathcal{N}_\gamma := \mathcal{N}_1 \cup \{X_a, X_b\} \cup \mathcal{N}_0$ denotes the crossings in γ. Furthermore, let $\mathcal{E}^\gamma_D$ denote the edges of D which are contained in γ; $\mathcal{E}^0_G$ denotes the edges of D which border $R_0 \cup R_{n+1}$.

For this next part, we will reference Figure 32 for clarity, but it will be obvious that our claims are completely general. We will refer to the edges of Figure 32 by their vertices, e.g., the broken edge from X_1 to X_8 is $\epsilon_{1,8}$.

Suppose a state σ is contributing. By assumption, its value on $\epsilon_{1,8}$ is 0, and by Lemma 7.5, $\sigma(1,0) = [\sigma(\epsilon_{1,8}) - \sigma(\epsilon_{1,9})] = 0$. Therefore we must have $\sigma(\epsilon_{1,9}) = 0$. We can then continue on in this fashion around the boundary of the knot, i.e., in general, the values of σ on $\mathcal{E}^0_D$ do no depend on σ. Specifically, we have $\sigma(e_1) = \sigma(h_1) = 0$, $\sigma(f_n) = 0$, and $\sigma(g_n) = -\mathrm{sgn}(n)$. Plugging these in, we obtain

$$\langle D|\sigma\rangle_1 = \frac{Nq^{-\mathrm{sgn}(1)/2-\sigma(g_1)}}{\left(q^{\mathrm{sgn}(1)}\right)_{\mathrm{sgn}(1)\sigma(g_1)}\left(q^{-\mathrm{sgn}(1)}\right)_{-\mathrm{sgn}(1)\sigma(g_1)-1}} = q^{-\mathrm{sgn}(1)/2-\sigma(g_1)} \tag{35}$$

$$\langle D|\sigma\rangle_n = \frac{Nq^{-\mathrm{sgn}(n)/2+\sigma(e_n)}}{\left(q^{-\mathrm{sgn}(n)}\right)_{\mathrm{sgn}(n)\sigma(e_n)}\left(q^{\mathrm{sgn}(n)}\right)_{-\mathrm{sgn}(n)\sigma(e_n)-1}} = q^{-\mathrm{sgn}(n)/2+\sigma(e_n)} \tag{36}$$

(in fact, note that $\sigma(e_n)$ and $\sigma(g_1)$ are determined as well, though this will not be needed). Now, let Ψ_σ be the set of states which agree with σ, except on the broken arc γ, i.e.

$$\Psi_\sigma = \{\rho \in \dot{\Psi} | \rho(\epsilon) = \sigma(\epsilon) \text{ if } \epsilon \not\subset \mathcal{E}^\gamma_D\}$$

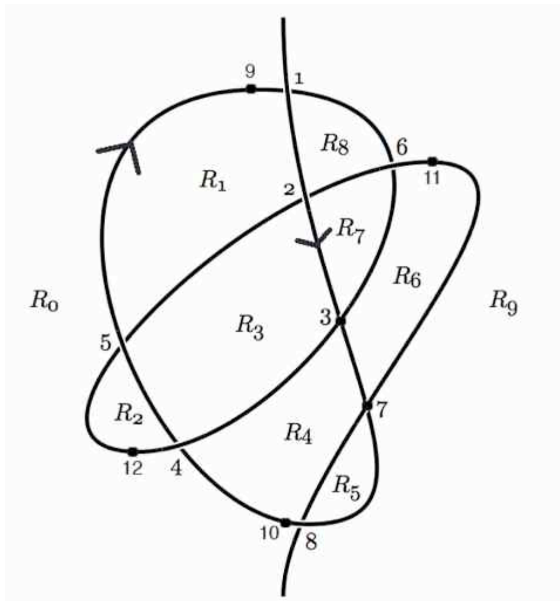

Figure 32:

Let us consider the sum

$$\sum_{\rho\in\Psi_\sigma}\prod_{\nu\in\gamma}\langle D|\rho\rangle_\nu$$

In our example this is

$$\sum_{\rho\in\Psi_\sigma}\langle D|\rho\rangle_1\langle D|\rho\rangle_2\langle D|\rho\rangle_3\langle D|\rho\rangle_7\langle D|\rho\rangle_8 \tag{37}$$

We have already calculated $\langle D|\rho\rangle_1 = q^{-\mathrm{sgn}(1)/2}q^{-\rho(g_1)}$. Then we have $g_1 = \epsilon_{1,2} = f_2$ (both crossings X_1 and X_2 are positive). Thus the term

$$\sum q^{-\rho(g_1)}\langle D|\rho\rangle_2 = \sum q^{-\rho(f_2)}S^{\rho(e_2)\rho(f_2)}_{\rho(g_2)\rho(h_2)} \tag{38}$$

is one of the sums from Lemma 7.7 (the second, to be precise) (recall that $\rho(e_2 = \epsilon_{2,5})$ and $\rho(h_2 = \epsilon_{2,6})$ are fixed, as the states we're summing over all agree off of γ). Furthermore, contributing states all have $\rho(f_2)\in[\rho(e_2),\rho(h_2)]$ anyway, so (38) becomes $\delta_{\rho(e_2),\rho(h_2)}q^{-1-\rho(h_2)}$ by Lemma 7.7. If there were any more crossings on γ between p and X_a, we would be able to repeat this previous computation exactly; the process stops at X_a. We then do the same thing in the other direction, stopping at X_b. In general, we get the expression

$$\sum_{\rho\in\Psi_\sigma}\prod_{\nu\in\mathcal{N}_\gamma}\langle D|\rho\rangle_\nu = \left(\sum_{\rho(e_a)\in\mathcal{G}}\langle D|\rho\rangle_a\cdot q^{-\rho(e_a)}\prod_{\nu\in\mathcal{N}_1}q^{-\mathrm{sgn}(\nu)/2}\delta_{\sigma(e_\nu),\sigma(h_\nu)}\right) \tag{39}$$

$$\times\left(\sum_{\rho(g_b)\in\mathcal{G}}\langle D|\rho\rangle_b\cdot q^{-\rho(g_b)}\prod_{\nu\in\mathcal{N}_0}q^{-\mathrm{sgn}(\nu)/2}\delta_{\sigma(f_\nu),\sigma(g_\nu)+\mathrm{sgn}(\nu)}\right) \tag{40}$$

which, using Lemma 7.8 and (34), becomes

$$Nq^{-\mathrm{sgn}(a)/2-\sigma(g_a)} \prod_{\mu\in\mathcal{N}_a^\star} \frac{1}{\left(q^{\mathrm{sgn}(a,\mu)}\right)_{\mathrm{sgn}(a,\mu)\sigma(a,\mu)}} \cdot \prod_{\nu\in\mathcal{N}_1} q^{-\mathrm{sgn}(\nu)/2}\delta_{\sigma(e_\nu),\sigma(h_\nu)} \tag{41}$$

$$\times\ Nq^{-\mathrm{sgn}(b)/2-\sigma(e_b)} \prod_{\mu\in\mathcal{N}_b^\star} \frac{1}{\left(q^{\mathrm{sgn}(b,\mu)}\right)_{\mathrm{sgn}(b,\mu)\sigma(b,\mu)}} \cdot \prod_{\nu\in\mathcal{N}_0} q^{-\mathrm{sgn}(\nu)/2}\delta_{\sigma(f_\nu),\sigma(g_\nu)+\mathrm{sgn}(\nu)} \tag{42}$$

Call a contributing state σ a *weight* if $\sigma(\epsilon)=0$ for $\epsilon\in\mathcal{E}_D^\gamma-\{X_a,X_b\}$, and if

$$\prod_{\nu\in\mathcal{N}_1}\delta_{\sigma(e_\nu),\sigma(h_\nu)}\prod_{\nu\in\mathcal{N}_0}\delta_{\sigma(f_\nu),\sigma(g_\nu)+\mathrm{sgn}(\nu)}\neq 0$$

Let Ω be the set of weights of D. Then we have

Theorem 7.8. *$F'_{N-1}(L)$ can be written as*

$$\pm N^{n-2}\prod_{\xi\in\mathcal{X}}\prod_{\nu\in\mathcal{N}} q^{-sgn(\nu)/2}\sum_{\sigma\in\Omega}\prod_{\nu\in\mathcal{N}}\left(q^{\sigma(e_\nu)-\sigma(g_\nu)}\prod_{\mu\in\mathcal{N}_\nu^\star}\frac{1}{\left(q^{sgn(\nu,\mu)}\right)_{sgn(\nu,\mu)\sigma(\nu,\mu)}}\right)$$

Proof. We have

$$\sum_{\sigma\in\dot\Psi}\prod_{\nu\in\mathcal{N}}\langle D|\sigma\rangle_\nu=\sum_{\sigma\in\Omega}\sum_{\rho\in\dot\Psi_\sigma}\prod_{\nu\in\mathcal{N}}\langle D|\sigma\rangle_\nu=\sum_{\sigma\in\Omega}\left(\prod_{\nu\notin\mathcal{N}_\gamma}\langle D|\sigma\rangle_\nu\cdot\sum_{\rho\in\dot\Psi_\sigma}\prod_{\nu\in\mathcal{N}_\gamma}\langle D|\rho\rangle_\nu\right)$$

So the theorem follows from (34) and (41).

□

The sum from Theorem 7.8 contains only the quantum factorials corresponding to non-degenerating tetrahedra, and so this proves Theorem 7.1.

Acknowledgements

I would like to give immeasurable thanks, firstly, to my parents, my sister, and my grandparents. Their unconditional support, admiration and engagement has meant more to me than I've probably ever realized. I especially appreciate the excellent cooking and coffee during the home stretch.

I want to thank my friends, especially from the math department. You have inspired me and taught me, and blurred the line between problem sets and entertainment. (Fab seven: thank you for reminding me that there is a line).

Above all, I would like to thank my advisor, Peter Kronheimer. Only in retrospect have I realized how your constant support and encouragement over the last three years are responsible not just for my present mathematical interests and direction, but for my continued interest in mathematics at all. You have engaged me in an inescapably knotted world which, try my best, I've yet to be able to evade. Neuroscience can wait; this thesis is dedicated to mathematics. (Thanks must then be due to Sylvain Cappell, for introducing me to knots in the first place).

Special thanks are due to Dylan Thurston, for several very valuable conversations.

References

[1] R. Benedetti and C. Petronio. *Lectures on Hyperbolic Geometry*. Springer, 1992.

[2] V. Chari and A. Pressley. *A Guide to Quantum Groups*. Cambridge University Press, 1994.

[3] VG Drinfeld. On almost cocommutative Hopf algebras. *Leningrad Math. J*, 1(2):321–342, 1990.

[4] VG Drinfeld. Quantum groups, from: Proceedings of the International Congress of Mathematicians, Berkeley, Calif. 1986. *Amer. Math. Soc. Providence, RI*, pages 798–820, 1987.

[5] K Hikami. Hyperbolic structure arising from a knot invariant. *Arxiv preprint math-ph*, Jan 2001.

[6] K Hikami. Asymptotics of the colored jones polynomial and the a-polynomial. *Nuclear Physics*, Jan 2007.

[7] R Kashaev. Quantum dilogarithm as a 6j-symbol. *Arxiv preprint hep-th*, Jan 1994.

[8] R Kashaev. A link invariant from quantum dilogarithm. *Modern Physics Letters A [Particles and Fields; Gravitation; ...*, Jan 1995.

[9] R Kashaev. The hyperbolic volume of knots from the quantum dilogarithm. *Letters in Mathematical Physics*, Jan 1997.

[10] C. Kassel. *Quantum Groups*. Springer, 1995.

[11] C. Kassel, M. Rosso, and V. Turaev. Quantum groups and knot invariants. *Panoramas et Syntheses*, 1997.

[12] R Kirby and P Melvin. The 3-manifold invariants of witten and reshetikhin-turaev for sl (2, c). *Inventiones Mathematicae*, Jan 1991.

[13] L.I. Korogodski and Y.S. Soibelman. *Algebras of Functions on Quantum Groups*. American Mathematical Society, 1998.

[14] S. Mac Lane. *Categories for the Working Mathematician*. Springer, 1998.

[15] S. Majid. *A Quantum Groups Primer*. Cambridge University Press, 2002.

[16] K. Matsuzaki and M. Taniguchi. *Hyperbolic Manifolds and Kleinian Groups*. Oxford University Press, 1998.

[17] Hitoshi Murakami and Jun Murakami. The colored jones polynomials and the simplicial volume of a knot. *arXiv*, math.GT, Jun 1999.

[18] Hitoshi Murakami, Jun Murakami, M Okamoto, and T Takata. Kashaev's conjecture and the chern-simons invariants of knots and links. *Experiment. Math*, Jan 2002.

[19] W.D. Neumann and D. Zagier. Volumes of hyperbolic three-manifolds. *Topology*, 24(3):307–332, 1985.

[20] J Ratcliffe. Foundations of hyperbolic manifolds. *books.google.com*, Jan 1994.

[21] NY Reshetikhin and VG Turaev. Ribbon graphs and their invaraints derived from quantum groups. *Communications in Mathematical Physics*, 127(1):1–26, 1990.

[22] Dylan Thurston. Hyperbolic volume and the jones polynomial: Grenoble notes. page 21, Nov 2004.

[23] W.P. Thurston. Geometry and Topology of Three-Manifolds. Princeton lecture notes, 1979. *Revised version*, 1991.

[24] V Turaev. The yang-baxter equation and invariants of links. *New Developments in the Theory of Knots*, Jan 1990.

[25] V Turaev. Quantum invariants of links and 3-valent graphs in 3-manifolds. *Publications Mathématiques de L'IHÉS*, Jan 1993.

[26] V.G. Turaev. *Quantum Invariants of Knots and 3-Manifolds*. Walter de Gruyter, 1994.

[27] Jeffrey R Weeks. Computation of hyperbolic structures in knot theory. *arXiv*, math.GT, Sep 2003.

[28] E Witten. Quantum field theory and the jones polynomial. *Communications in Mathematical Physics*, Jan 1989.

[29] Yoshiyuki Yokota. On the volume conjecture for hyperbolic knots. *Arxiv preprint math.QA*, Jan 2000.

[30] Yoshiyuki Yokota. On the potential functions for the hyperbolic structures of a knot complement. *Geometry & Topology Monographs*, Jan 2002.