ISSN 1464-8997 (on line) 1464-8989 (printed)

Geometry & Topology Monographs
Volume 4: Invariants of knots and 3-manifolds (Kyoto 2001)
Pages 337–362

The algebra of knotted trivalent graphs and Turaev's shadow world

Dylan P. Thurston

Abstract Knotted trivalent graphs (KTGs) form a rich algebra with a few simple operations: connected sum, unzip, and bubbling. With these operations, KTGs are generated by the unknotted tetrahedron and Möbius strips. Many previously known representations of knots, including knot diagrams and non-associative tangles, can be turned into KTG presentations in a natural way.

Often two sequences of KTG operations produce the same output on all inputs. These "elementary" relations can be subtle: for instance, there is a planar algebra of KTGs with a distinguished cycle. Studying these relations naturally leads us to Turaev's *shadow surfaces*, a combinatorial representation of 3-manifolds based on simple 2-spines of 4-manifolds. We consider the knotted trivalent graphs as the boundary of a such a simple spine of the 4-ball, and to consider a Morse-theoretic sweepout of the spine as a "movie" of the knotted graph as it evolves according to the KTG operations. For every KTG presentation of a knot we can construct such a movie. Two sequences of KTG operations that yield the same surface are topologically equivalent, although the converse is not quite true.

AMS Classification 57M25; 57M20, 57Q40

Keywords Knotted trivalent graphs, shadow surfaces, spines, simple 2-polyhedra, graph operations

1 Introduction

In this paper we study the algebra of *knotted trivalent graphs* (KTGs). A knotted trivalent graph is a framed[1] embedding of a trivalent graph into $\mathbb{R}^3$, modulo isotopy. These KTGs support some simple operations, forming an algebra-like structure. Every knot may be presented as a sequence of KTG operations starting with elementary graphs. Thus we may use KTGs as a novel representation of knot theory via generators and relations.

---

[1] See Section 2 for the precise notion of framing we use.

Published 3 February 2004:

We may compare a KTG presentation with other representations of knots, such as:

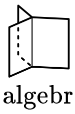
- Planar knot diagrams;
- Braid closures;
- n-bridge representations;
- Pretzel representations;
- Rational tangles and algebraic knots [5];
- Parenthesized tangles [2, 11]; and
- Curves in a book with three pages[2] [7].

These representations of knots all take an inherently 3-dimensional object (a knot) and squash it into 2 dimensions (as in a knot diagram) or sometimes even into 1 dimension (as in a parenthesized tangle). The algebra of KTGs deals more directly with knots as 3-dimensional objects, while strictly generalizing all of the above representations of knots, in the sense that, for instance, any knot diagram can be turned into a sequence of operations on KTGs in a natural way.[3] We will illustrate how KTGs generalize knot diagrams in Section 4.

The next step in studying the algebra of KTGs is to find the relations in the algebra, and, in particular, the *elementary* relations (Section 5): pairs of sequences of operations that produce the same output on all inputs. Further justification for calling these relations "elementary" comes from the fact that in other spaces that support the same operations the elementary relations are automatically satisfied, while other relations give us non-trivial equations to solve. Tracing out the track of the KTG as it evolves through a sequence of operations, we construct a *movie surface* (Section 6), a decorated simple 2-polyhedron. If two different sequences of operations generate the same movie surface, then they are universally equal. The converse is not quite true.

In fact, movie surfaces are a special case of *shadow diagrams*, as we briefly discuss in Section 7. Shadow diagrams for 3-manifolds and for 4-manifolds bounded by 3-manifolds were introduced by Turaev. They were initially introduced to describe links inside circle bundles over a surface [15]. The construction was later generalized to allow descriptions of all 3-manifolds [14, 16]. From a quantum topology point of view, shadow diagrams encapsulate the algebra of

---

[2] A book with three pages looks like this: 

[3] Here "algebra" is used in the universal algebra sense of a set supporting some operations

quantum 6j-symbols in a concise way; more classically, a shadow diagram of a 3-manifold can be thought of as an analogue of a pair of pants decomposition of a surface.

In summary, every KTG presentation of a knot gives a certain abstract surface representing that knot complement. Thus it turns out that our intrinsically 3-dimensional representation of knot theory can also be encoded in 2-dimensional terms. The relation between the various spaces and constructions is summarized in Figure 1.

$$\begin{tikzcd}
\text{Framed links} \arrow[r, "\text{exterior}"] & \text{Framed link exteriors} \arrow[r, "\text{surgery}"] & \text{Closed 3-manifolds} \\
\begin{array}{l}\text{Sequences of}\\ \text{KTG operations}\end{array} \arrow[u, "\text{presentation}"] \arrow[r, "\begin{array}{l}\text{time}\\ \text{evolution}\end{array}"'] & \begin{array}{l}\text{Collapsible}\\ \text{shadow surfaces}\end{array} \arrow[u, "\begin{array}{l}\text{by making}\\ S^1\text{-bundle}\end{array}"'] \arrow[r, "\text{cap off } \partial"'] & \text{Closed shadow surfaces} \arrow[u, "\begin{array}{l}\text{by making}\\ S^1\text{-bundle}\end{array}"']
\end{tikzcd}$$

Figure 1: A summary of the relations between links, 3-manifolds, KTGs and shadow surfaces

This paper is an exposition of results that were, at least implicitly, previously known. Rather, our aim is to give an exposition of the relationship between shadow surfaces and knotted trivalent graphs. In a future paper, we will prove some theorems that came out of this work, including the relationship between hyperbolic volume and minimal complexity shadow diagrams representing a knot. In addition, the unifying framework of KTGs is part of an ongoing project to find new combinatorial presentations of knots, which may be more algebraically manageable than the full-blown algebra of KTGs. See Section 8 for more on both of these.

1.1 Acknowledgements

A very large part of the credit for this work must go to Dror Bar-Natan, with whom I have had a long, productive, and fun collaboration. It was a great experience coming up with the algebra of KTGs with him and puzzling over the wide variety of mysterious relations that we found together. In this papers, Sections 2 through 5 are joint work with him. In addition, I would like to thank Riccardo Benedetti, Francesco Costantino, Robion Kirby, Tomotada Ohtsuki, Carlo Petronio, A. Referee, Dale Rolfsen, Chung-chieh Shan, Vladimir Turaev, and Genevieve Walsh for a great many helpful conversations and comments.

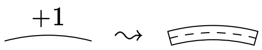
This work was supported by an NSF Postdoctoral Research Fellowship, a JSPS Postdoctoral Research Fellowship, and by BSF grant #1998119.

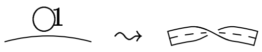
2 The space of knotted trivalent graphs

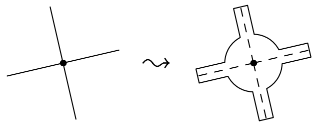
There are relatively few operations on knots. Traditional operations, like connect sum, cabling, or general satellites, when applied to non-trivial knots, never yield hyperbolic knots, but almost all knots (by any reasonable measure of complexity) are hyperbolic. Thus typical knot operations are no good for breaking the vast majority of knots down into simpler pieces. To fix this, we will pass to the larger space of knotted trivalent graphs. This space will allow more operations; enough to generate all knots from a few simple generators.

“Graphs” in this paper might more properly be called “1-dimensional complexes”; that is, they may have multiple edges, self loops, and circle components.

Definition 1 A *framed graph* or *fat graph* is a thickening of an ordinary graph into a surface (not necessarily oriented): the vertices are turned into disks and the edges are turned into bands attaching to the disks. More abstractly, a framed graph is a 1-dimensional simplicial complex Γ together with an embedding $\Gamma \hookrightarrow \Sigma$ of Γ into a surface Σ as a spine; more combinatorially, a framed graph is a graph with a cyclic ordering on the edges incident to each vertex and 1 or -1 on each edge (representing a straight or flipped connection, respectively), modulo reversing the ordering at a vertex and negating the elements on the adjoining edges.







The notion of spines comes from PL topology: a *spine* of a simplicial complex Y is a subcomplex X of Y onto which Y collapses, where collapsing means successively removing pairs of a k-simplex Δ^k and a $(k+1)$-simplex Δ^{k+1}, where Δ^{k+1} is the unique $(k+1)$-simplex having Δ^k on its boundary.

For instance, the two framed graphs with spine $\Gamma = S^1$ are the annulus and the Möbius band.

Definition 2 A *knotted trivalent graph* (KTG) is a trivalent framed graph Γ embedded (as a surface) into $\mathbb{R}^3$, considered up to isotopy.

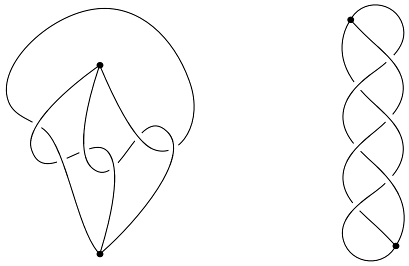


Figure 2: Two knotted theta graphs, one of which is unknotted. Which one is it?

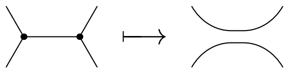
Definition 3 An *unknotted KTG* is a planar KTG: a KTG which factors through an embedding of S^2 in $\mathbb{R}^3$.

In diagrams, KTGs will be drawn with the blackboard framing unless explicitly indicated otherwise.

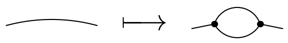
3 Elementary operations

The advantage of KTGs over knots or links is that they support many operations. By an *operation* we mean a function from input trivalent graphs to an output graph, depending only on combinatorial choices (and no topological data).

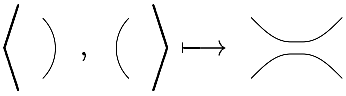
- *Unzip* takes an edge which connects two distinct vertices and splits it in two, as though you were unzipping a zipper along the edge. It reduces the number of vertices by 2.

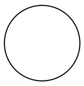
- *Bubbling* adds a small loop along an edge of the KTG. It increases the number of vertices by 2.

- *Connect sum* takes two input trivalent graphs, with a chosen edge on each one (and a chosen side of each edge), and splices the two edges together. Note that this operation has two, independent, inputs (else the operation depends on a choice of an arc and so is not well-defined in the sense above). Bubbling is equivalent to connect sum with an unknotted theta graph .



- The identity for connect sum is the *unknot.*



Unzip and bubbling operate on one input graph, while connect sum operates on two inputs and the unknot has no inputs.

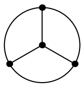
Exercise 4 Check that these operations are well-defined. In particular, connect sum is well-defined for the same reason that the connect sum of two knots is well-defined. (Hint: you can shrink one of the trivalent graphs into a small ball and slide it along the other.)

Note that "operations" on knot diagrams such as changing a crossing or self connect sum are not operations in our sense, since they depend not only on combinatorial data, but also on a topological choice of *how* to perform the move.

Also, we only consider these operations in the forward direction. The reverse operations cannot always be performed, and so are not KTG operations, in that they do not act on the space of all KTGs with a given underlying trivalent graph.

These operations do not suffice to construct interesting knots; for instance, every KTG constructed with these operations will be planar (unknotted). In addition to these operations, we will use three elementary KTGs as generators:

- The unknotted *tetrahedron.*



- The two minimally twisted *Möbius bands*, with a positive (resp. negative) half twist. There is no blackboard framing for these Möbius bands, so the half twist is indicated by a little local picture of the surface.

Remark The distinction between operations on KTGs and generators for the algebra of KTGs is somewhat a matter of taste. For instance, the unknot can be constructed by unzipping the tetrahedron twice, so need not be included in the algebra; however, just as the unit of a ring is generally considered to be part of the structure of a ring, it is more natural to consider the unknot as a part of the algebraic structure of KTGs.

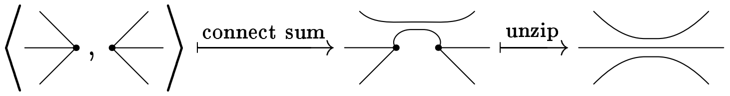
4 Constructing knots

The KTG operations we have defined are quite powerful. In particular, they generate knot theory.

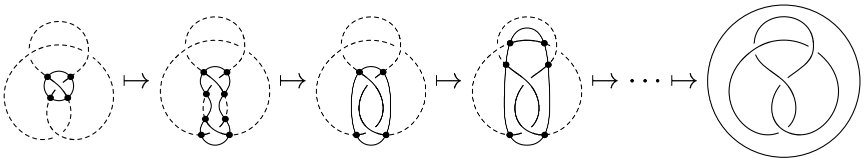
Theorem 1 *Any KTG can be obtained from the unknotted (planar) tetrahedron and the two minimally twisted Möbius bands using unzip, connect sum, bubbling, and the unknot.*

Each of the presentations of knots listed at the beginning of the paper can be turned into a sequence of KTG operations and thus gives a proof of Theorem 1. In the proof of the theorem below we will show how to turn a knot diagram into a sequence of KTG operations. But first, let us define some *composite operations* formed by composing the elementary operations.

- A connect sum along an edge followed by an unzip along one of the two newly created edges gives a *vertex connect sum*: take a vertex in each of two KTGs and join the incoming edges pairwise. For a given matching of the edges, there are several ways of performing this operation as a sequence of elementary operations, all yielding identical results.



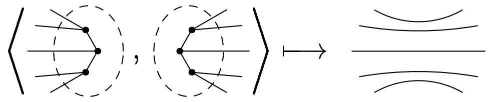
- More generally, there is a notion of *tree connect sum*: for any KTGs K_1, K_2, and isomorphic open subsets T_1, T_2 of the skeletons of K_1, K_2 which are homeomorphic to trees, we can join T_1 and T_2 by a connect sum and repeated unzips. Topologically, you can think of straightening out the two subtrees into some standard embedding in a ball (which you can always do, since any embedding of a tree in $\mathbb{R}^3$ is trivial) and then removing the balls from both K_1 and K_2 and identifying the two

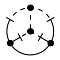


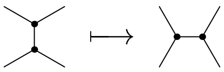
(1) (2) (3) (4) (5)

Figure 3: Generating the figure 8 knot with KTG operations by sweeping a circle outwards.

sphere boundary components, connecting corresponding points on the boundaries.



- Connected sum with Möbius bands can *change the framing* on an edge by any desired (integral or half integral) amount.
- Tree connect sum with the tetrahedron  along the dashed subtree performs the *Whitehead move.*



To prove Theorem 1, we use a sweep-out argument based on a knot diagram. Mark a circle around a small piece of a knot diagram and consider the KTG formed by the circle and the portion of the knot diagram on the inside. If only one feature of the knot diagram is enclosed, there are only a few possibilities for the resulting KTG and it is straightforward to check that they may all be generated with our generators and operations. Then sweep this marked circle outwards by stages, performing a few elementary KTG operations at each stage.

Rather than giving an exhaustive proof, we will work through the steps in a sweepout of a figure 8 knot in Figure 3, which provide a good sampling of the cases. At each stage, the marked circle and the portion of the knot inside it are drawn solid.

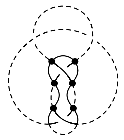
(1) By changing framings, we may turn the unknotted tetrahedron into the crossed tetrahedron . This is what we obtain by drawing our initial marked circle around a crossing, as on the left of Figure 3.

(2) At the next step, we push the circle across another crossing. We can achieve this by taking connect sum with a second crossed tetrahedron.

(3) Pushing the circle across a maximum of a strand (looking outwards from the strand) is achieved by an unzip move. Two of these happen in the next step.

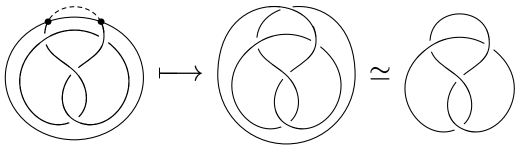
(4) Pushing the circle across a minimum is a bubbling move.

(5) We proceed in this way, using one crossed tetrahedron per crossing of our diagram, until we have pushed the circle all the way out.



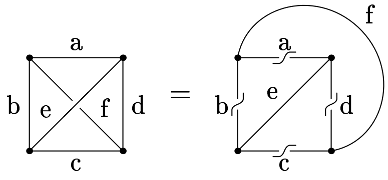
We have succeeded in generating our knot, plus a trivial unknot component. There are a few ways to avoid this extra component. For instance, in the intermediate steps, one might simply not include the portions of the circle that touch the exterior region. This version of step (2) in Figure 3 shown at the right.

Alternatively, we could modify the very last unzip move before we push the marked circle completely off the knot diagram.

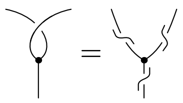


With a similar procedure, we can generate any KTG. We end up using at most one tetrahedron per trivalent vertex and one tetrahedron per crossing in our diagram.

Exercise 5 Show that the crossed tetrahedron is equal to an unknotted tetrahedron with the following framing changes:



Hint: start by showing the more elementary equality:



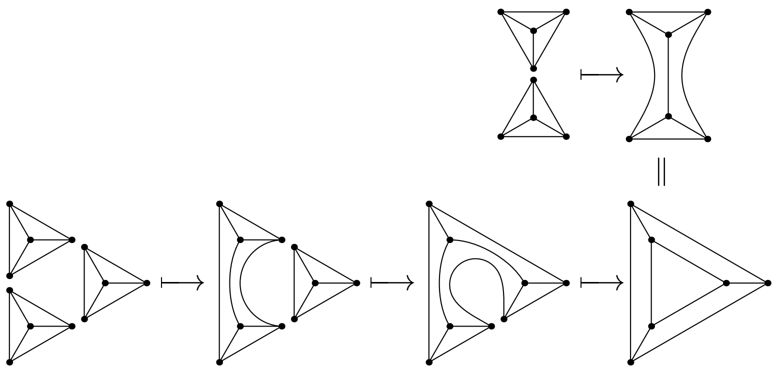


Figure 4: The pentagon relation: an unknotted triangular prism can be made in two ways. It may be obtained by the vertex connect sum of two tetrahedra (top); alternatively, it may be obtained by the vertex connect sum of three tetrahedra followed by an unzip (bottom).

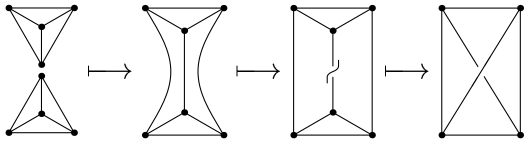


Figure 5: The hexagon relation: unzipping a triangular prism (made from two tetrahedra) in a twisted fashion yields a twisted tetrahedron.

5 Relations and elementary relations

Knot theory is not freely generated by the tetrahedron and Möbius strip; there are some relations. Two of them are the pentagon (Figure 4) and the hexagon (Figure 5).

The names come from the "pentagon" and "hexagon" relations in the theory of non-associative tangles [2], which become the above relations when non-associative tangles are interpreted as sequences of KTG operations in a natural way. In non-associative tangles, the pentagon and hexagon are the essential identities, the ones that require significant work to solve; there are, however, many more identities which are "elementary", in that they are automatically satisfied in the framework. The same thing is true in the theory of knotted trivalent graphs. This leads us to the central question of this paper, which we can phrase in several different ways.

Question 6 What are the elementary relations in the algebra of KTGs? Which composite KTG operations produce the same output for all inputs? Which KTG operations are associative? What relations in the algebra of KTGs are automatically true in other natural spaces supporting the trivalent graph operations (unzip, bubbling, connect sum, unknot)? What is the operad[4] of KTG operations?

Here are some examples of elementary equalities. Each elementary equality equates two composite KTG operations that take the same number and types of KTGs as input.

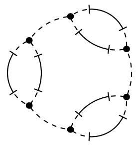
- Any two operations performed on disjoint pieces of a knotted graph commute with each other.
- The connect sum operation on knots is commutative and associative.
- The parallel operation on links (replacing a knot component with a 2 components parallel with respect to the framing) is cocommutative and coassociative.
- The operations of vertex connect sum and tree connect sum are well-defined: they do not depend on which of the matching edges we connect sum and the order in which we perform the unzips.
- The left hand ("3") side of the pentagon equation above does not depend on the choices. We took 3 tetrahedra and did 2 vertex connect sums followed by an unzip. The resulting triangular prism has a symmetry that the construction did not have. This is not an accident. In fact, if we take *any* three knotted tetrahedra (with an identification of the underlying graphs with the underlying graph of the three standard tetrahedra above) and do the same sequence of operations (vertex connect sums between two pairs followed by an unzip) in the three possible ways, the result depends only on the original knotting of the tetrahedra and not on the order of operations.
- The algorithm for turning a knot diagram into a sequence of KTG operations depended on a choice of how to sweep out the knot diagram by circles. In fact, the result is independent of the sweep out, even if the crossed tetrahedra  are replaced by *arbitrary* knotted tetrahedra.

Some of these identities towards the end of the list may surprise you. They can all be given good topological explanations. For instance, the description of tree connect sum in terms of gluing balls is clearly independent of choices. The case

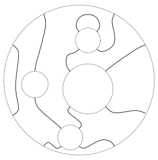
---

[4] To be precise, this is a typed operad, whose types are the framed trivalent graphs.

of the "3" side of the pentagon can be reduced to some tree connect sums using the following exercise.



Exercise 7 Express the combination of three tetrahedra as on the "3" side of the pentagon equation as a tree-connect sum with the graph at right along the dashed subtrees: one subtree is connected to each of the input tetrahedra. In other terms, we can straighten out the subtrees being acted on in each of the three input tetrahedra before performing the operation. The symmetry of the operation then becomes obvious.



The "knot diagram" case is more involved. In the case that the outer circle of the small tetrahedron is unknotted this operation amounts to gluing together tangles with 4 legs following the pattern of the knot diagram; however, the operation is well-defined in general. The general statement is that knotted trivalent graphs with a distinguished cycle form a planar algebra in a sense similar to that of Vaughan Jones [9]. (This structure has also been called a "spider" [10] or a "spherical category" [3].) Briefly, for every arc diagram as on the right, there is a well-defined operation that takes as input KTGs corresponding to the interior cycles (with a distinguished cycle containing a corresponding set of vertices) and produces as output a KTG with a distinguished cycle. The proof would take us too far afield, but it follows from more general statements below.

6 Movie surfaces

To give some order to the zoo of seemingly "elementary" equivalences between composite knotted trivalent graph operations, we add an extra dimension. Consider making a movie of the graph as it evolves in S^3. The graph traces out a surface with some simple singularities in a 4-dimensional space. We can continue this surface across each of the elementary operations and generators to form a continuous singular surface, as illustrated in Table 1. For instance, to do an unzip move along an edge, first shrink the edge to zero length and then split the two new branches apart. The resulting movie is shown in the first line of the table.

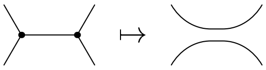
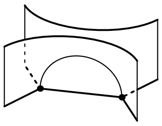
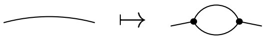
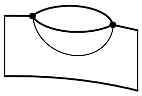
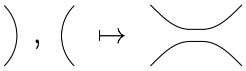
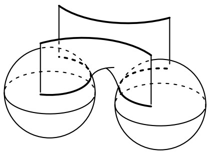
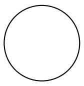
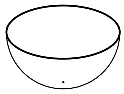
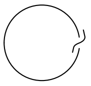
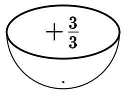
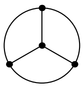
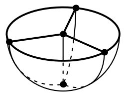
Name	Operation	Movie surface
Unzip		
Bubbling		
Connect sum		
Unknot		
Möbius band		
Tetrahedron		

Table 1: The surface created from a movie of the knotted trivalent graph operations and generators. The surfaces depicted all lie in a local 3-dimensional slice of a 4-dimensional space; although the pictures appear 3-dimensional, they are really pictures of surfaces in 4 dimensions.

Some comments on this table:

- These are surfaces in a 4-dimensional space (the evolving S^3), although the depicted portions lie inside a 3-dimensional slice. This is related to the fact that our knotted graph operations locally lie in a plane, as does the unknotted tetrahedron (but not the Möbius strip; see below).
- The unzip and bubbling surfaces lie in $S^3 \times I$, with one input S^3 (at the bottom of the picture) and one output S^3 (at the top). Outside of the portion depicted, there is no change to the graph, so the corresponding surface is a product.
- The connect sum operation has two input S^3's, depicted as two spheres in the picture. (The interiors of these spheres are not included in the ambient 4-manifold.) As before, the output S^3 is at the top of the picture. This surface lies inside a thrice-perforated S^4.

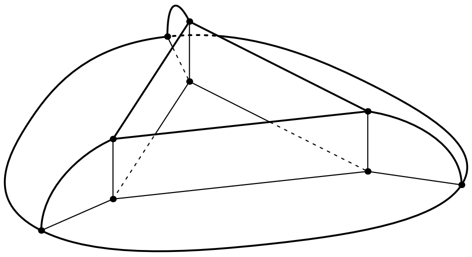


Figure 6: The movie surface for the "3" side of the pentagon. The picture lies in B^4, with the edges drawn with thick lines lying on the boundary of B^4.

- The unknot, Möbius strip, and unknotted tetrahedron have no inputs. The corresponding surfaces lie in B^4.
- The Möbius strip is the only one of these operations or generators that is not essentially planar. The underlying unframed knotted graph of the Möbius strip is a simple loop, so it bounds a disk in B^4, just like the unknot. To distinguish it from the unknot we must somehow record the framing. This will be explained more in Section 7, but for now we will just attach the framing information to the surface itself, as indicated in the diagram. (The convention in diagrams is that this framing information, the "gleams", may be attached to the surface at multiple points; each surface component has a total gleam, which is the sum of the gleams attached at various points. In particular, if no gleams are indicated on a surface component the total gleam is 0.)

For a somewhat more involved example of the surface constructed by the moving knotted trivalent graph, Figure 6 shows the surface associated to the "3" side of the pentagon. Observe that the surface has the symmetries of the triangular prism; that is, it has the symmetries that were missing from the previous description with KTG operations.

Exercise 8 Convince yourself that Figure 6 is the movie surface representing the sequence of operations on the bottom line of Figure 4. The action in the movie is from the inside out: start with small neighborhoods of each vertex (corresponding to 3 unknotted tetrahedra) and push the neighborhoods outwards until they meet in pairs (performing the vertex connect sum moves) and then encompass the remaining edge (performing the unzip move).

The movie surface for the corresponding operation on 3 knotted tetrahedra (which takes the tetrahedra and connects them according to the same pattern)

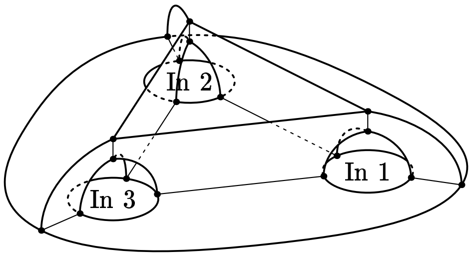
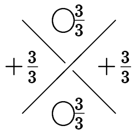


Figure 7: The movie surface for the operation which takes three knotted tetrahedra and attaches them in the same pattern as on the "3" side of pentagon.

is shown in Figure 7. This surface lives inside B^4 with three balls removed. The three input components of the boundary are labelled "In 1", "In 2", and "In 3"; each of these input components lie on one of the removed balls. The picture suggests that the input tetrahedra are unknotted, but in fact this surface can be embedded (uniquely!) in B^4 minus 3 balls so as to match any 3 given input knotted tetrahedra.

More generally, consider the movie surface obtained from the sequence of KTG operations used to construct a knot (plus an extra unknotted component) from a knot diagram in Section 4. Tracing out the surface through the sequence of KTG operations, we find that the movie surface is constructed by taking a disk and attaching an annulus along the curve of the knot diagram, with the gleams



on the regions of the knot diagram around each vertex. Each complementary region to the knot diagram gets assigned several gleams, one for each incident vertex; add up the gleams for each vertex. (These gleams come from Exercise 5). If we forget the gleams, the surface is the mapping cylinder of the map from S^1 to the disk specified by the knot diagram. The resulting surface coincides with Turaev's "shadow cylinder" on the knot. See Figure 8 for the example of the figure 8 knot. Note that this is a picture of abstract surface which does not embed in $\mathbb{R}^3$; the transverse intersections of surfaces which appear in the diagram are only artifacts of the immersion. There are two boundary components of this surface (including the external square), corresponding to the fact that we constructed a 2-component link (with one unknotted component) in Figure 3.

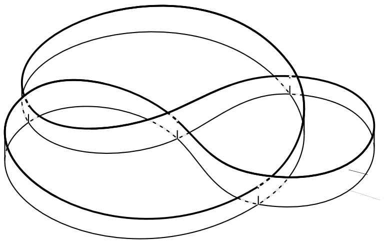


Figure 8: The movie surface corresponding to the sequence of KTG operations used to construct the figure 8 knot in Figure 3.

Observation 9 *When we pass from a sequence of KTG operations to the associated surface, we often see additional symmetries. We do not get the same surface for two topologically different sequences of KTG operations.*

Thus the movie surface seems to be a good representation of a sequence of KTG operations. This observation will be justified below, where we will see that a KTG may be reconstructed from its movie surface.

Exercise 10 Check that each of the sequences of equivalent KTG operations listed in Section 5 yield homeomorphic surfaces, with one exception: coassociativity of the parallel operation.

7 Shadow Surfaces

The movie surfaces constructed in the previous section are, in fact, special cases of *shadow surfaces*, which we will now introduce. The key observation to make about each of the drawings in Table 1 is that the ambient 4-manifold collapses onto the union of the movie surface and the S^3 input(s). To see this collapsibility, first note that outside a neighborhood of the operation, the movie surface is constant in time, and so the 4-manifold away from the movie surface may be collapsed onto the input. Furthermore, the operations all take place inside a 3-dimensional slice; the 4-manifold outside of the 3-dimensional slice can again be collapsed. So we are left with showing that, for each surface Σ in Table 1 considered as a 3-dimensional surface, $\mathbb{R}^3$ collapses down onto Σ. But the complementary regions of each Σ are all 3-balls touching the output surface in just one connected 2-balls, so we can collapse each 3-ball, as desired.

Note that the argument above does not apply if, e.g., we turned the unzip surface upside down (one complementary region touches the output in two places), turned the bubble surface upside down (one complementary region does not touch the output), or considered a self-connect sum with only one input (one complementary region is not a 3-cell); corresponding to the fact that none of these operations are topologically well-defined inside S^3. (Shadow surfaces do give a topological interpretation to each of these surfaces, but the result is a KTG in some 3-manifold, not generally S^3.)

A somewhat stronger statement is that in each case the ambient 4-manifold is diffeomorphic to a collar over the S^3 input(s), together with a regular neighborhood of Σ in a 4-manifold, with the surface embedded in a locally flat way. (Without this caveat, the surface might locally be the cone over some non-trivial knot in S^3.)

The situation is somewhat easier to think about in the case when there are no inputs, so we have a KTG presentation of a knotted graph Γ. In this case, all of B^4 collapses down onto the movie surface Σ and, to reconstruct the pair (B^4, Γ) from Σ, we just need to describe how to "thicken" Σ into its regular neighborhood. This thickening should take a surface (with singularities) and yield a smooth 4-manifold. Away from the singularities, we take a disk bundle over the surface, and continue in a natural way over the singularities. If we are only interested in 3-manifold topology, we may consider the circle bundle at the boundary of the disk bundle.

To understand this situation a little better, let us consider the situation one dimension down: instead of constructing 3-manifolds from singular surfaces, let us construct 2-manifolds from graphs. (See Figure 1 or Table 2 for a guide to the analogy.)

	Dimension 1 (graphs)	Dimension 2 (simple surfaces)
codim 1	Fat graph	Simple spine of 3-manifold
codim 2	3-dimensional handlebody	Shadow of slim 4-manifold
∂ of codim 2	Pair-of-pants decomposition	Shadow of 3-manifold

Table 2: An analogy between shadows of 3-manifolds and pair-of-pants decompositions of surfaces: the regular neighborhoods of polyhedra of varying dimension and codimension.

We have already seen one way to construct 2-manifolds from graphs, in Definition 1: The construction of fat (or framed) graphs from a trivalent graph Γ.

This construction is nearly equivalent to triangulations of closed surfaces: if we glue a disk onto each circle component of the boundary of a fat graph, we get a closed surface Σ with Γ drawn on it. Assuming that Γ has no circle components, the dual cell division to Γ is a triangulation of Σ, in the weak sense (for instance, two sides of a triangle may be glued to each other).

But this construction is in codimension 1 (a graph is thickened into a surface), while the construction we are interested in is in codimension 2 (a surface is thickened into a 4-manifold). For the proper analogue, we should consider a codimension 2 thickening of a graph: that is, thickening a graph into a 3-manifold. A regular neighborhood of a graph Γ in a 3-manifold is a handlebody; its boundary is a closed surface Σ. If we cut Σ along circles surrounding each edge of Γ, we end up with a thrice-perforated sphere for each vertex of Γ. This is also known as a pair-of-pants decomposition of Σ. (We are again assuming that Γ has no circle components.)

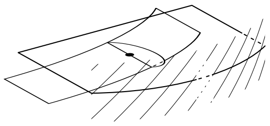
Moving up one dimension, let us consider the thickening of surfaces. First let us be precise about the type of singular surfaces we consider.

Definition 11 A *simple polyhedron* or *simple surface* is a 2-dimensional simplicial complex in which the link of every point is locally homeomorphic to a neighborhood of some point in the (closed) cone over the 1-skeleton of a tetrahedron. The *boundary* of a simple surface is the subset of the surface locally homeomorphic to a point on the 1-skeleton of the tetrahedron itself.

The possible local models for the interior of a simple surface are $\mathbb{R}^2$, the three page book, and the open cone over a tetrahedron. The possible local models for the boundary are the upper half plane and the upper half of the three page book. All the surfaces considered in this paper are simple polyhedra.

In codimension 1, if a manifold M^3 is a regular neighborhood of a simple surface Σ^2, then Σ is called a *spine* of M^3. Every 3-manifold with non-empty boundary has a spine. Spines of 3-manifolds, like fat graphs, are closely related to triangulations: the dual cell complex to a triangulation of a closed 3-manifold M^3 is a spine for M^3 minus a neighborhood of each vertex of the triangulation. The principal advantage of spines over triangulations is that it is easier to consider various kinds of degeneracy: the dual to a triangulation is a *standard spine*, in which all of the edges of the spine are intervals connecting two vertices and all of the faces are disks. In general, a *standard surface* is a simple surface satisfying this extra condition that the edges be intervals and the faces be disks. It turns out that the thickening of a standard surface into an orientable 3-manifold is completely determined by the surface, if it exists [4].

In codimension 2, we may consider a regular neighborhood W^4 of a 2-polyhedron Σ^2 in an orientable 4-manifold. For reasonable topological control, we require that Σ be locally flat at generic points of Σ, so that W is generically a disk bundle over Σ. Suppose for the moment that Σ is a smoothly embedded closed ordinary surface. Then the disk bundle is determined by its Euler number.[5] This number is what we called the "gleam" in Section 6 in the discussion of the Möbius strip.



In general, the normal bundle to each smooth component Y of Σ is a 2-plane bundle on Y. In order to assign a well-defined Euler number to this bundle, we need to define a trivialization on the boundary. At generic points on the boundary of Y, there are two other regions of Σ attached to the 1-skeleton. At vertices on the boundary of Y, there is also an opposite region attached, but it may be ignored: around each boundary component of Y, there is a unique way to map an annulus or a Möbius strip S (the hashed region in the diagram to the right) in to Σ so that the boundary of Y is a spine of S, and Y is otherwise disjoint from S.

After homotopy, we can assume that S is normal to Y; it then provides two vectors in the normal bundle to Y over each point in ∂Y. In the case that S is an annulus, we can pick one of the two vectors and get a trivialization of the normal bundle over the boundary, and so we can define a relative Euler number relative to Σ. In case S is a Möbius strip, we get a half-trivialization lying between two honest trivializations (each obtained by inserting a half twist in S); we correspondingly define the relative Euler number to be a half-integer. It turns out that these Euler numbers, the gleams, are enough to reconstruct W from Σ [14, 16]. (Recall that W is a regular neighborhood of Σ.) Let us state this precisely.

Definition 12 A *shadow surface* is a simple polyhedron (possibly with boundary) together with a gleam in $\mathbb{Z}$ (resp. $\mathbb{Z}+\frac{1}{2}$) on each face which has an even (resp. odd) number of Möbius strip components around the boundary.

Definition 13 A (strict) *shadow* of a 4-manifold W^4 is a closed shadow surface Σ^2 which is a spine of W, with the gleam on each face Y of Σ equal to the Euler class of the normal bundle to Y.

---

[5] Σ itself need not be oriented: a disk bundle over a surface has a well-defined Euler number if the total space is orientable.

A *(relative) shadow* of a pair (W^4, Γ) of a 4-manifold W^4 with a (framed) trivalent graph Γ embedded in ∂W is a simple polyhedron Σ^2 properly embedded in W so that $\partial\Sigma = \Gamma$ and W collapses onto Σ. Recall that the framing really means the thickening of Γ into a surface; attach W to the original graph running along the center of this surface. The resulting slightly larger surface can be used to give gleams to all the regions of W in a uniform way.

Remark Our terminology differs somewhat from that of Turaev [14, 16]. Turaev defines shadows to be equivalence classes of surfaces modulo some moves (including the pentagon and hexagon in Figures 9 and 10). Since these equivalence classes turn out to be nearly the same as 4-manifolds in the most interesting case, we prefer to reserve the name "shadow" or "shadow surface" for the actual decorated surface, rather than the equivalence class.

Proposition 14 (Turaev) *Every shadow surface Σ is the shadow of a 4-manifold W, which is unique up to PL homeomorphism preserving Σ.*

Remark In the setting of the above proposition, W may admit non-trivial self-homeomorphisms preserving Σ.

Proposition 15 *The movie surface associated to a KTG presentation of a knotted graph Γ is a collapsible shadow for (B^4, Γ). Furthermore, every collapsible shadow surface arises in this way.*

Proposition 16 *Every collapsible shadow surface Σ (with arbitrary gleams) is the shadow of (B^4, Γ) for some KTG Γ.*

Proof Let Σ be a shadow for (W^4, Γ). By hypothesis, W collapses onto Σ and Σ collapses to a point, so W collapses to a point and is therefore B^4. $\square$

Remark A KTG presentation is equivalent to its movie surface Σ together with the horizontal slicing given by time. This gives a particular type of Morse function on Σ, one which exhibits the fact that Σ is collapsible. (Precisely, for $t_0 \leq t_1$, $\Sigma_{t_0 \leq t \leq t_1} \cup C(\Sigma_{t_0})$ should be collapsible, where $C(\Sigma_{t_0})$ is the cone on Σ_{t_0}.)

Question 17 Does every collapsible simple 2-polyhedron Σ admit a Morse function exhibiting the fact that Σ is collapsible?

Definition 18 A *shadow representation* of an oriented 3-manifold M^3 is a 4-manifold W^4 with $\partial W = M$, together with a shadow for W.

Every oriented 3-manifold has a shadow representation: The effect of attaching a disk to the boundary of a shadow surface is surgery on the corresponding knot. Since every link in S^3 has a shadow representation (since it has a KTG presentation) and every 3-manifold is surgery on a link, we can get shadow representations of any 3-manifold.

On the other hand, not every 4-manifold has a shadow representation. A shadow representation of a 4-manifold W provides a handle decomposition of W with only 0-, 1-, and 2-handles, so we must in particular have $H^3(W) = H^4(W) = 0$. This class of 4-manifolds is an interesting one; for instance, they are the handlebodies that are homeomorphic to Stein domains [8].

Question 19 Are any two standard shadow surfaces representing the same 4-manifold are related by a sequence of pentagon (Figure 9), hexagon (Figure 10), and 2-0 (Figure 11) moves and their inverses?

The answer to Question 19 is probably "no". There is a well-known conjecture which is closely related to the case when the shadow is contractible:

Conjecture 20 (Andrews-Curtis) *Any two contractible simple polyhedra can be related by the pentagon, hexagon, and 2-0 moves and their inverses, ignoring all gleams.*

There is also a generalized Andrews-Curtis conjecture, which drops the hypothesis that the polyhedra be contractible. The usual form of the Andrews-Curtis conjecture is in terms of balanced presentations of the trivial group, but this version is equivalent. This conjecture is generally believed to be false: there are many known potential counterexamples. While Conjecture 20 and Question 19 are not directly related (in either direction), many potential counter-examples to the Andrews-Curtis conjecture can be turned into potential counter-examples to Question 19 above. For instance, the examples of Akbulut and Kirby [1] are of this type. Question 19 can be thought of as an embedded version of the Andrews-Curtis conjecture.

If we want a shadow calculus for 3-manifolds as opposed to 4-manifolds, we need to allow for some move that changes the 4-manifold. One such move, which may be used to achieve either the Kirby I move or the Fenn-Rourke move, is shown in Figure 12.

In joint work with Francesco Costantino, we have shown that this is sufficient.

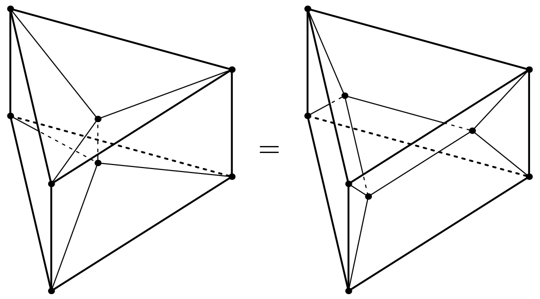


Figure 9: The pentagon relation in the shadow world.

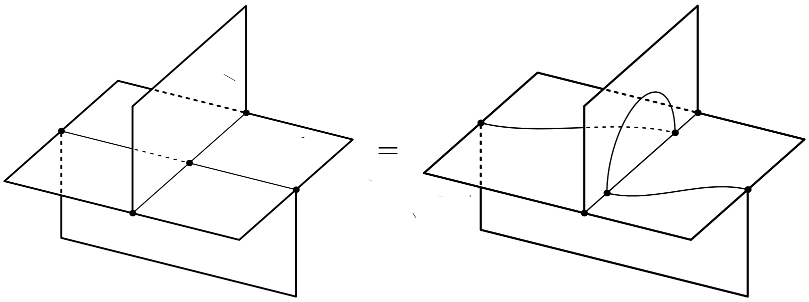


Figure 10: The hexagon relation in the shadow world. On the right hand side, the lower rectangle attaches along the path that runs on the upper rectangle; the result is not embeddable in $\mathbb{R}^3$.

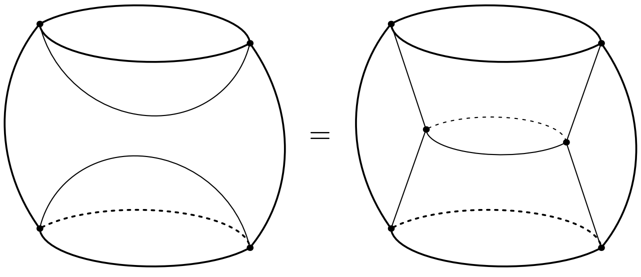


Figure 11: The 2-0 move, an additional move needed to be able to change the topology of regions of the shadow.

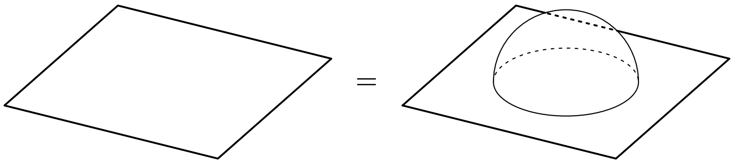


Figure 12: +1 surgery on an unknot is trivial, as seen in the shadow world.

Theorem 2 (Costantino-Thurston, to appear) *Any two simply-connected shadow presentations of the same 3-manifold are related by the pentagon, hexagon, 2-0, and* ± 1*-bubble moves.*

Notice that adding a bubble, and so increasing the second homology of the surface, is enough to avoid Andrews-Curtis problems.

8 What's it good for?

Besides providing a unified framework for viewing many different types of knot representations, the algebra of KTGs as presented in this paper has many possible applications, to be explored more in future papers.

8.1 Representations of the algebra of KTGs

As mentioned above, there are other natural spaces that support the same set of KTG operations, with the same elementary relations. We may call these "TG-algebras". A TG-algebra consists of a space associated to each (abstract) trivalent graph, and maps between them corresponding to the elementary operations. To find a knot invariant with values in a TG-algebra it suffices to find values for the tetrahedron and Möbius strips which satisfy the pentagon and hexagon relations.

This construction is known for some TG-algebras. There is one TG-algebra related to representations of a group G; the quantum invariants of Reshetikhin and Turaev come from representations on this space. For some other spaces, this is apparently new. For instance, the space of chord diagrams on the graph, the target for the universal Vassiliev invariant, supports these same operations; in these terms, there is an elegant characterization of the Kontsevich integral.

Proposition 21 *The Kontsevich integral Z is the unique universal Vassiliev invariant which has an extension to knotted trivalent graphs well-behaved under the TG operations, and for which a half-framing change acts by multiplication by* $\exp(\theta/4)$.

Although the knot invariant is unique, the extension to KTGs is not unique, but the ambiguity is understood. The condition on the framing change is just a normalization condition. This is a reformulation of previously known results [6, 13, 12], but it has extra symmetry.

Another TG-algebra which is a potential target for a KTG invariant is more-or-less dual to the space of representations mentioned above: the space associated to a KTG consists of measures on G^E/G^V: a copy of the group G for each edge, modulo an action for each vertex. One reason to be interested in representations on this space is that it provides a natural setting for trying to understand Witten's asymptotics conjectures on asymptotic behaviour of quantum invariants.

8.2 Finding new specializations

The algebra of knotted trivalent graphs may be too large for some purposes. We have only described the operations (up to equivalence) implicitly, by equivalence of the underlying surfaces. However, there are several special cases which are easier to deal with, forming, for instance, an ordinary algebra with a single associative multiplication. Some of these special cases were mentioned at the beginning of this paper. The KTG point of view suggests several more special cases which are worth investigating, including a category of *annular braids*.

8.3 Complexity issues

Every knot diagram with n crossings can be turned into a KTG presentation with n or fewer tetrahedra. The converse is not true: KTG presentations may be much more efficient than knot diagrams. For instance, it is easy to construct a sequence of links where the linking number grows exponentially in the number of KTG operations.

For bounds in the other direction, recall that the Gromov norm of a knot complement is (up to a constant) equal to the sums of the hyperbolic volumes of the hyperbolic pieces of the geometric decomposition of the complement. (In particular, for an iterated torus knot the Gromov norm is 0, and for a hyperbolic knot the Gromov norm is essentially the volume.) It can be shown that the Gromov norm is less than a constant times the number of tetrahedra in any KTG presentation. So the minimum number of tetrahedra in a KTG representation of a knot is bounded above by the Gromov norm, and below by the usual crossing number of the knot.

It is an open question where in between these bounds the minimum number of tetrahedra lies. A related result is that any geometric 3-manifold has a shadow presentation with a number of vertices at most quadratic in the Gromov norm. This is joint work with Francesco Costantino; a paper is in preparation.

References

[1] **Selman Akbulut**, **Robion Kirby**, *A potential smooth counterexample in dimension* 4 *to the Poincaré conjecture, the Schoenflies conjecture, and the Andrews-Curtis conjecture*, Topology 24 (1985) 375–390

[2] **Dror Bar-Natan**, *Non-associative tangles*, from: "Geometric topology (Athens, GA, 1993)", AMS/IP Stud. Adv. Math. 2 (1997) 139–183

[3] **John W Barrett**, **Bruce W Westbury**, *Spherical Categories*, Adv. Math. 143 (1999) 357–375

[4] **B G Casler**, *An imbedding theorem for connected* 3*-manifolds with boundary*, Proc. Amer. Math. Soc 16 (1965) 559–566

[5] **John H Conway**, *An enumeration of knots and links, and some of their algebraic properties*, from: "Computational Problems in Abstract Algebra (Proc. Conf., Oxford, 1967)", Pergamon, Oxford (1970) 329–358

[6] **Vladimir G Drinfel'd**, *On Quasitriangualar Quasi-Hopf Algebras and a Group Closely Connected with* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, Leningrad Math. J. 2 (1991) 829–860

[7] **Ivan A Dynnikov**, *Three-page representation of links*, Uspekhi Mat. Nauk 53 (1998) 237–238. English translation: Russian Math. Surveys 53 (1998) 1091–1092

[8] **Robert E Gompf**, *Handlebody construction of Stein surfaces*, Ann. of Math. 148 (1998) 619–693

[9] **Vaughan F R Jones**, *Planar Algebras, I*, Technical report, U. C. Berkeley (1999), to appear in New Zealand J. Math., `arXiv:math.QA/9909027`

[10] **Greg Kuperberg**, *Spiders for Rank 2 Lie Algebras*, Comm. Math. Phys. 180 (1996) 109–151, `arXiv:q-alg/9712003`

[11] **Thang T Q Le**, **Jun Murakami**, *The universal Vassiliev-Kontsevich invariant for framed oriented links*, Compositio Math. 102 (1996) 41–64, `arXiv:hep-th/9401016`

[12] **Christine Lescop**, *About the Uniqueness of the Kontsevich Integral*, Journal of Knot Theory and its Ramifications 11 (2002) 759–780, `arXiv:math.GT/0004094`

[13] **Sylvain Poirier**, *The Configuration Space Integral for Links and Tangles in* $\mathbb{R}^3$, Ph.D. thesis, Fourier Institute, University of Grenoble (2000), `arXiv:math.GT/0005085v2`

[14] **Vladimir G Turaev**, *Topology of shadows* (1991), preprint

[15] **Vladimir G Turaev**, *Shadow links and face models of statistical mechanics*, J. Differential Geom. 36 (1992) 35–74

[16] **Vladimir G Turaev**, *Quantum invariants of knots and 3-manifolds*, de Gruyter Studies in Mathematics 18, W de Gruyter, Berlin (1994)

Department of Mathematics, Harvard University
Cambridge, MA 02138, USA

Email: `dpt@math.harvard.edu`

Received: 25 November 2003 Revised: 24 January 2004