# Fixing the functoriality of Khovanov homology

DAVID CLARK
SCOTT MORRISON
KEVIN WALKER

We describe a modification of Khovanov homology [13], in the spirit of Bar-Natan [2], which makes the theory properly functorial with respect to link cobordisms.

This requires introducing "disorientations" in the category of smoothings and abstract cobordisms between them used in Bar-Natan's definition. Disorientations have "seams" separating oppositely oriented regions, coming with a preferred normal direction. The seams satisfy certain relations (just as the underlying cobordisms satisfy relations such as the neck cutting relation).

We construct explicit chain maps for the various Reidemeister moves, then prove that the compositions of chain maps associated to each side of each of Carter, Rieger and Saito's movie moves [8; 7] always agree. These calculations are greatly simplified by following arguments due to Bar-Natan and Khovanov, which ensure that the two compositions must agree, up to a sign. We set up this argument in our context by proving a result about duality in Khovanov homology, generalising previous results about mirror images of knots to a "local" result about tangles. Along the way, we reproduce Jacobsson's sign table [10] for the original "unoriented theory", with a few disagreements.

# 1 Introduction

Khovanov homology [13; 14] (see also Bar-Natan [2]) is a "categorified" invariant: it assigns to a link a graded module (or a complex of such) rather than a "scalar" object such as a number or a polynomial. Thus we expect not merely a module for each link, but also a functor which assigns module isomorphisms to each isotopy between links. (This isomorphism should depend only on the isotopy class of the isotopy.) That is, given two links and a specific isotopy between them, we want an explicit isomorphism between their Khovanov invariants, not merely the knowledge that the Khovanov invariants are isomorphic. Unfortunately, the original unoriented version of Khovanov homology gives slightly less than this – the isomorphisms assigned to isotopies are well-defined only up to sign.
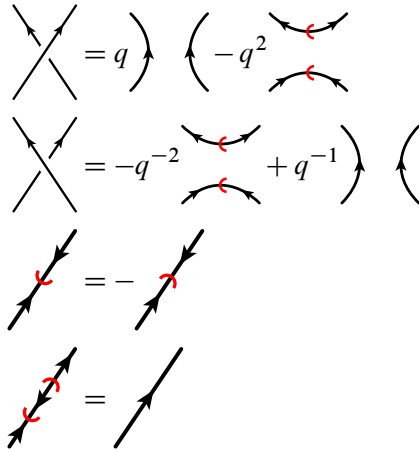
Unoriented Khovanov homology also gives more: the functor extends to surface cobordisms in $B^3 \times I$ (but still with a sign ambiguity) by Jacobsson [10]. More precisely, let $\mathcal{L}$ be the above category of oriented links and (isotopy classes of) isotopies between them, and let $\mathcal{C}$ be the category whose morphisms are (isotopy classes of) oriented surfaces properly embedded in $B^3 \times I$. If we associate to each isotopy between links the track of the isotopy in $B^3 \times I$, we get a functor $\mathcal{L} \hookrightarrow \mathcal{C}$, and the Kh functor on $\mathcal{L}$ is the pullback of an extended Kh functor on $\mathcal{C}$. The extended Kh also has a sign ambiguity.

The aim of this paper is to fix the above sign issues.

For motivation, consider the "precategorified" situation. Unoriented Khovanov homology is based on the unoriented Kauffman bracket polynomial, with skein relation shown in Figure 1 (with a further writhe correction, which introduces a dependence on the orientations of the link). Closely related is the quantum $\mathfrak{su}_2$ polynomial, which has a skein theory based on piecewise oriented (or "disoriented") tangles, as shown in Figure 2 (see Kirby and Melvin [18]). The two polynomials (and their associated TQFTs) differ only by a sprinkling of signs. The Kauffman bracket has the advantage of simpler (unoriented) objects and trivial Frobenius–Schur indicators, while the quantum $\mathfrak{su}_2$ polynomial has the advantage of producing positive-definite TQFTs (that is, TQFTs with nicer signs).



Figure 1: A version of the Kauffman skein relations

Our strategy is to categorify the disoriented skein relation of the quantum $\mathfrak{su}_2$ polynomial, rather than the unoriented Kauffman skein relation. We introduce the appropriate category of disoriented surface cobordisms, and then imitate Bar-Natan's approach. We find that disorientations also lead to nicer signs in the categorified setting:

**Theorem 1.1** *There is a functor* Kh *from the category of oriented links in $B^3$ and (isotopy classes of) isotopies between them to the category whose objects are graded complexes of disoriented smoothings and abstract disoriented cobordisms between smoothings (modulo local relations) and whose morphisms are graded chain isomorphisms (modulo chain homotopy).*

Figure 2: The "disoriented" $\mathfrak{su}_2$ skein relations

**Theorem 1.2** *The functor* Kh *extends to the category of oriented links in* $B^3$ *and oriented surface cobordisms (modulo isotopy) in* $B^3 \times I$.

We split the statement into two theorems because functoriality with respect to isotopies of links would be expected of any link invariant taking values in a category, while functoriality with respect to surface cobordisms is a special feature of Khovanov homology.

For each link $L$, Hom $(\mathrm{Kh}(\varnothing), \mathrm{Kh}(L))$ is a doubly graded vector space (the second grading is the homological shift of a chain map).[1] It follows from Theorem 4.1 that this vector space is isomorphic to the one constructed from the unoriented theory as in Bar-Natan [2]. Hence it is isomorphic to Khovanov's original construction, and in particular its graded Euler characteristic is the Jones polynomial of $L$.

Similar results have been obtained independently by Caprau [6].

In [16], Khovanov and Rozansky define a link homology theory which is a categorification of the generalized Jones polynomial based on $U_q(\mathfrak{sl}_n)$. We do not know whether our invariant is equivalent to Khovanov and Rozansky's for $n = 2$, but it seems reasonable to conjecture that it is.

---

[1]Here, Hom $(\mathrm{Kh}(\varnothing), \mathrm{Kh}(L))$ is the space of chain maps up to homotopy between $\mathrm{Kh}(\varnothing)$ and $\mathrm{Kh}(L)$ in the category of complexes disoriented cobordisms. Equivalently, we could apply the functor Hom $(\varnothing, \text{—})$ to each object and morphism in the complex $\mathrm{Kh}(L)$, and then take the homology of the resulting complex of graded vector spaces.

We actually get much more than a functor on cobordisms. We can construct a 4–category (or, if you prefer, a 4–dimensional version of a planar algebra) whose 3–morphisms are tangles in $B^3$ and whose 4–morphisms are elements of appropriate Khovanov homology modules. This 4–category enjoys the following duality or "Frobenius reciprocity" type property:

**Theorem 1.3** *Given compatible oriented tangles $P$, $Q$ and $R$, there is a duality isomorphism between the spaces of chain maps up to homotopy*

$$F\colon \mathrm{Hom}_{\mathrm{Kh}}\left(\llbracket P \bullet Q \rrbracket, \llbracket R \rrbracket\right) \xrightarrow{\cong} \mathrm{Hom}_{\mathrm{Kh}}\left(\llbracket P \rrbracket, \llbracket R \bullet \bar{Q} \rrbracket\right).$$

The notation $P \bullet Q$ indicates a "horizontal composition" of tangles $P$ and $Q$. See Section 3.1 for a more precise statement.

The duality isomorphisms are coherent in the following sense (although this is not proved in this paper). To each such isomorphism we can associate an isotopy of links in $S^3$ – roughly speaking we slide $Q$ from the bottom of $S^3$ to the top. Then two composable sequences of duality isomorphisms give the same result if the associated isotopies in $S^3 \times I$ are isotopic.

The paper is organized as follows.

Section 2 defines the invariant. We introduce the appropriate category of disoriented cobordisms, associate a chain complex based on this category to each oriented planar tangle diagram, and associate a morphism of complexes to each Reidemeister and Morse move.

Section 3 verifies that our construction is well-defined. We show that if two different sequences of Reidemeister and Morse moves are related by movie moves, then the associated morphisms of chain complexes are equal. Along the way, we prove the first part of the above duality result (Theorem 1.3).

Section 4, as its title suggests, contains miscellaneous results. We show that setting $\omega = 1$ in our construction recovers the signs from [10]. We show that modulo signs, our invariant agrees with the original unoriented version. We give an example calculation, showing that in the new construction, the cobordisms which "attach a handle to a strand" on either side of a crossing give homotopic chain maps, whereas the old construction gave maps homotopic only with a sign. Finally, we discuss the possibility of extending the invariant from oriented tangles to disoriented tangles.

# 2   The new construction

## 2.1   Disorientations

In this paper we follow the Bar-Natan approach of defining Khovanov homology in terms of surface cobordism categories — categories whose objects are (possibly crossingless) tangles in $B^3$ and whose morphisms are surface cobordisms between tangles. We'll deal with three sorts of tangles and surfaces: unoriented (and possibly nonorientable), oriented, and disoriented. We assume reader is familiar with the former two categories.

A disoriented 1– or 2–manifold is a piecewise oriented manifold where each component of the interface between differently oriented domains is equipped with a preferred normal direction. In figures, we indicate this normal direction with a fringe pointing in the preferred direction. We'll call the interface between differently oriented domains a disorientation seam.

We almost always (and usually without comment) consider disoriented surfaces modulo the local fringe relations illustrated in Figure 3. If $\omega$ is a primitive fourth root of unity ($\omega^2 = -1$), we will see below that we get a version of Khovanov homology that satisfies functoriality. If $\omega = 1$, then we reproduce the original unoriented version of Khovanov homology, simply because the disorientations become irrelevant. (We keep track of factors of $\omega$ explicitly, rather than just writing $\omega = i$ everywhere, so that we can do calculations in both the old and the new setup in parallel.)
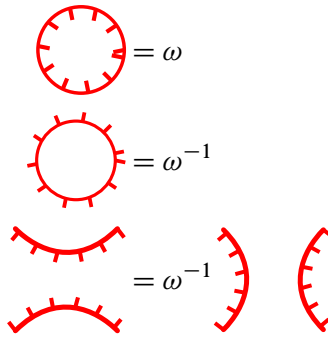
Figure 3: Disorientation relations

## 2.2 Cobordism categories

The main goal of this paper is to construct a functor from **OrTang**, the category of oriented tangles and oriented cobordisms in $B^4$, to Kom(**DisAb**), a category of chain complexes based on abstract disoriented cobordisms between disoriented crossingless planar diagrams. Along the way we'll meet several other variant cobordism categories. In this subsection we introduce the various categories we'll need. The categories will be given compound names like **OrTang**, Kom(**DisAb**) and Kom(**UnAb**); we'll start by explaining the meanings of the components of the names.

The manifolds in the categories (1–manifolds for objects, 2–manifolds for morphisms) can be unoriented, oriented or disoriented, which we denote by **Un**, **Or** and **Dis**. In all cases, we think of the objects as 1–manifolds embedded in $B^2 \times I = B^3$, with specified endpoints along the circle $\partial B^2 \times \{\frac{1}{2}\} \subset \partial B^3$. In particular there are morphisms between two tangles $X$ and $Y$ only if their boundaries on the circle match, and in that case the morphism will be some 2–manifold with boundary $X \cup -Y \cup (\partial \times I)$.

We now introduce three categories of tangles. The first one, **Tang**, is the one of real interest; it denotes the category whose objects are arbitrary tangles in $B^3$ and whose morphisms are isotopy classes of surface cobordisms embedded in $B^3 \times I = B^4$.

The second, **PD**, should be thought of as a "combinatorial model" of **Tang**. The objects of **PD** are tangles in $B^3$ which are in general position with respect to the projection $p_z: B^3 \cong B^2 \times I \to B^2$. The morphisms of the category can be described by generators and relations. The generators are:

- isotopies through tangles in general position.
- Morse moves; birth or death of a circle, or a saddle move.
- Reidemeister moves.

One should think of these generators as those isotopies which have at most one "singular time slice"; that is, one moment at which the projection of the link to $B^2$ is not generic, and the only the simplest types of singularity are allowed to occur. These simplest singularities are, of course, simply the Morse and Reidemeister moves.

The first relation we impose is a boring one; composing an "isotopy through general position tangles" with any other morphism simply gives a morphism of the same type, given by gluing the isotopies together. We then impose more relations, the movie moves of Carter, Rieger, and Saito [8; 7] (see also Roseman [21]). The unoriented versions of these moves are shown in Figure 4 (thanks to Carter and Saito for originally drawing these diagrams!), using the numbering scheme introduced by Bar-Natan in [2]. Note that we also need to consider variations involving mirror images and/or crossing changes.
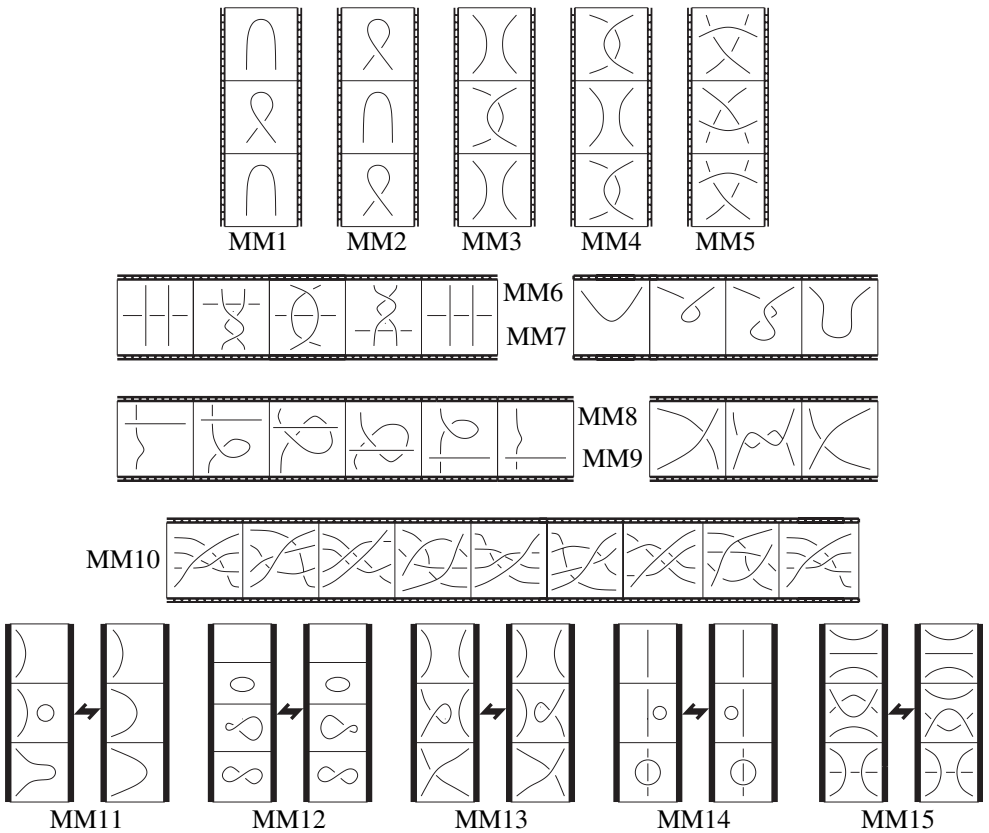


Figure 4: Carter and Saito's unoriented movie moves

They prove a theorem to the effect that two unoriented cobordisms between unoriented tangles represented by compositions of Morse and Reidemeister moves are isotopic if and only if those compositions are related by a sequence of movie moves. To describe the relations we impose in **OrPD**, we need the oriented version of this, which, by much the same argument as they gave, requires a separate version of each unoriented movie move for each possible orientation of the strands (subject to some constraints; movies involving saddles must have strands oriented appropriately so the saddles are valid morphisms).

Finally, note that in **DisPD** there are both additional Reidemeister moves (sliding a disorientation through a crossing) and additional movie moves, involving this new Reidemeister move. As we're not attempting here to extend Khovanov homology to all of **DisTang**, we'll omit most of the details of this, except what appears in Section 4.4.

Actually, we need to add a little more data to the objects in **PD**; a specified ordering on the crossings. (The chain complexes we eventually assign to diagrams will vary in boring but important ways according to the ordering of the crossings.) In addition to the morphisms described above (Reidemeister and Morse moves), we need to add "reordering morphisms", which are all isomorphisms. Further, we need to modify our notion of the Reidemeister moves so that the source and target tangles have (arbitrarily) ordered crossings – but all such different Reidemeister moves differ simply by pre- or post-composition with reordering isomorphisms.

Finally, **Ab** denotes a category whose objects are tangles without any crossings (think of them as embedded in $B^2 \times \{\frac{1}{2}\} \subset B^3$). The morphisms from a crossingless tangle $X$ to a crossingless tangle $Y$ are linear combinations of abstract (*not* embedded in $B^4$) surfaces, modulo the relations:

(2-1)
$$\left( \text{sphere} \right) = 0 \qquad \left( \text{torus} \right) = 2$$
$$\left( \text{cylinder} \right) = \frac{1}{2}\left( \left( \text{cap-torus} \right) \quad \left( | \right) + \left( | \right) \quad \left( \text{torus-cap} \right) \right)$$

The coefficients are in some ring containing $\frac{1}{2}$. Note that for **DisAb** these relations are imposed away from the disorientation seams. The last relation above is called the neck cutting relation. In **DisAb** we of course also impose the fringe relations (Figure 3, earlier). In **DisAb** it is unnecessary to set the 2–sphere equal to zero as it follows from the fringe relations that any connected, closed, orientable, disoriented surface whose Euler characteristic is not a multiple of 4 is equivalent to zero. Further it's easy to see via neck cutting that a genus $2n + 1$ connected closed oriented surface is equal to $2^{1-n}\left( \text{genus surface} \right)^n$, and that there are no unorientable disoriented surfaces.

As explained by Bar-Natan [2] and Naot [20], setting the genus three surface to zero in **UnAb** leads to the original version of Khovanov homology, while setting it to a nonzero complex number gives something isomorphic to Lee homology [19]. Although it makes very little difference for this paper, we'd like to encourage leaving this surface unevaluated, as described in [20]. This makes the morphism spaces into $\frac{\mathbb{Z}}{2}[\smile\!\smile\!\smile]$ modules. For convenience, we'll abbreviate $\frac{\mathbb{Z}}{2}[\smile\!\smile\!\smile]$ simply as $\mathcal{R}$; although for the purposes of the rest of the paper you can take $\mathcal{R}$ to be any ring with 2 invertible, if you prefer.

We'll now prove a lemma describing the morphism spaces in **DisAb**, in particular showing that the category is nontrivial and describing bases for some of the morphism spaces.

First, let's say that a disoriented circle has *total disorientation number zero* if the numbers of forward- and backward-facing disorientation marks are equal when we traverse the circle. (The total disorientation number itself is only defined up to a sign unless we independently have an orientation of the entire circle.) For such a circle, let's call the *standard disc* the disoriented disc in which the disorientation seams have been produced by successively connecting each pair of disorientation marks which face towards each other with no unconnected disorientation marks between them. Similarly the *standard punctured torus* is the disoriented punctured torus in which the disorientation seams have been produced in the same way. Finally, the *standard cylinder* is the disoriented cylinder, with boundary two copies of the same disoriented surface, in which the disorientation seams have been produced in this way near each boundary component separately. In particular, no disorientation seams connect one boundary component to the other. If the circle has $2m$ disorientation marks, the standard cylinder is equal to $\omega^m$ times the identity cylinder.

### Lemma 2.1

- *All closed surfaces in **DisAb** evaluate to a $\mathbb{Z}[\frac{1}{2}, \omega]$ multiple of a disjoint union of copies of $\smile\!\smile\!\smile$, and the surfaces $\bigcup_k \smile\!\smile\!\smile$ are all linearly independent. That is,*

$$\mathrm{Hom}_{\mathbf{DisAb}}(\varnothing, \varnothing) = \mathcal{R}[\omega].$$

- *If $D_1$ and $D_2$ are oriented diagrams in **DisAb** with the same boundary, and $D_1 \cup \overline{D_2}$ consists of $k$ oriented circles, then $\mathrm{Hom}_{\mathbf{DisAb}}(D_1, D_2)$ is a free $\mathcal{R}[\omega]$ module of rank $2^k$, spanned by surfaces consisting of discs or punctured tori attached to the boundary circles.*

- If $D_1$ and $D_2$ are arbitrary diagrams in **DisAb** with the same boundary, and $D_1 \cup \overline{D_2}$ consists of $k$ disoriented circles each with total disorientation number zero then $\mathrm{Hom}_{\textbf{DisAb}}(D_1, D_2)$ is again a free $\mathcal{R}[\omega]$ module of rank $2^k$, spanned by surfaces consisting of standard discs or standard punctured tori attached to the boundary circles.

*(In fact, the first two statements are just special cases of the third.)*

**Remark**  When the boundary circles don't have total disorientation number zero the morphism spaces are more complicated. For example, there are no morphisms between circles with different total disorientation numbers. Further, endomorphisms of the circle with total disorientation number two is a rank 2 module, rather than a rank 4 module as for the total disorientation number zero case. We won't need to know these details; only what appears here is needed for the proof of Theorem 4.1, although also see Section 4.4.

**Proof**  Usually such an argument about the basis for a theory defined by local relations proceeds by finding a functor (eg a TQFT) to vector spaces. We won't do this, but instead describe an "evaluation algorithm" which writes an arbitrary element of the morphism space as a linear combination of the indicated elements, and then show that modifying that arbitrary element by one of the local relations does not change the output of the algorithm. In particular, if some nontrivial linear combination of the proposed basis elements was actually equivalent to zero by applying the relations, then the algorithm would have to produce zero; we'll show that the algorithm leaves each of the proposed basis elements unchanged.

Given a surface in $\mathrm{Hom}_{\textbf{DisAb}}(D_1, D_2)$ with $D_1$ and $D_2$ satisfying the hypotheses of the lemma, the algorithm proceeds by:

(1) Replace a collar of each circle with $2m$ disorientation marks with $\omega^{-m}$ times the standard cylinder on that circle.

(2) Neck cut around the middle of each of those standard cylinders.

(3) At this point we have a disjoint union of standard discs and punctured tori, along with some disoriented closed surfaces. Apply the "disoriented neck cutting relation"



at each closed disorientation seam.

(4)   Now we have a disjoint union of standard discs and punctured tori, along with some *oriented* closed surfaces. Replace any even genus closed surface with $0$, and each closed surface of genus $2n+1$ with $2^{1-n} \langle\!\!\!\ominus\!\ominus\!\ominus\!\!\!\rangle^n$. (Notice this leaves $\langle\!\!\!\ominus\!\ominus\!\ominus\!\!\!\rangle$ unchanged.)

First observe that this algorithm leaves a standard disc or standard punctured torus unchanged (this requires a short calculation; it's good practice).

Next, we need to check each relation. Some of the relations are obvious; in particular the first two disorientation relations in Figure 3, as well as the sphere and torus relations from (2-1).

Neck cutting is not so hard. It's easy to see that if $Z$ is the result of neck cutting on $Y$, then the first three steps of the algorithm run in exactly the same way on each, and in particular do not modify $Z$ or $Y$ in the vicinity of the neck. After these three steps we have $Z'$ and $Y'$, where $Z'$ is obtained from $Y'$ by neck cutting along a circle on one of the oriented closed surfaces. It's easy to check that the fourth step of the algorithm gives the same result in either case.

Finally, we need to check the last disorientation relation in Figure 3. Suppose $Y$ and $Z$ are identical morphisms, except in some disc where they appear as the left and right sides of the relation. After the first two steps of the algorithm, we have $Y'$ and $Z'$, which are still identical outside this disc (so far we've only made modifications in a neighborhood of the boundary). Let's assume without loss of generality that in $Y'$ the two disorientation seams involved are part of two distinct circles, while in $Z'$ the two seams are part of the same circle. Consider then the pair of pants $P$ that includes the disc in which the relation takes place, and a neighborhood of the three disorientation circles. We thus have
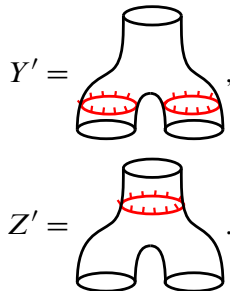
Continuing now with the third step of the algorithm, we obtain

$$Y'' = \frac{1}{4}\left(\omega^2 \;\; + \;\; + \;\; + \omega^{-2} \right)$$

and

$$Z'' = \frac{1}{2}\left( \;\; + \omega^{-2} \right).$$

Knowing by now that we can freely neck cut, we neck cut each of the cylinders appearing in the terms of $Z''$, and see that the algorithm produces exactly the same results.                                                                          □

In all of the above categories, we allow objects to carry an integer, thought of as a "formal grading shift", just as in [2]. We'll denote this grading shift by a power of $q$. We grade all of the morphism spaces, so that for a cobordism $C$ with source object $q^{m_1}D_1$ and target object $q^{m_2}D_2$, each with $k$ boundary points, $\deg(C) = \chi(C) - k/2 + m_2 - m_1$. It is not hard to see that these degrees are additive under both composition and planar operations (in fact, $\chi(C) - k/2$ and $m_2 - m_1$ are each additive separately). The local relations in Equation (2-1) are clearly degree homogeneous, so our grading makes sense on the quotient.

Given any category $\mathcal{C}$ with linear morphism spaces (called "preadditive" in [2]), we can form a category Mat$(\mathcal{C})$ whose objects are tuples of objects of $\mathcal{C}$ (written as formal direct sums), and whose morphisms are matrices of morphisms of $\mathcal{C}$. Composition is given by multiplying matrices.

As an example to illustrate the grading and matrix conventions, let us recall the "delooping" isomorphism described by Bar-Natan [3]. This is an isomorphism in Mat(**UnAb**) (there is an identical isomorphism in Mat(**DisAb**)) between $\bigcirc$ and $q\varnothing \oplus q^{-1}\varnothing$, given by the matrices

$$\begin{pmatrix} \\ \tfrac{1}{2} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \tfrac{1}{2} & \end{pmatrix}.$$

That these matrices are inverses follows immediately from the relations in Equation (2-1) (and a quick calculation that the double torus is zero, by neck cutting). Observe

that all the matrix entries here are degree 0 morphisms, once the grading shifts on the source and target objects have been taken into account.

We can also form the category $\mathrm{Kom}(\mathcal{C})$, whose objects are chain complexes built out of $\mathrm{Mat}(\mathcal{C})$, and whose morphisms are degree 0 chain maps modulo chain homotopy.

So, reviewing the nomenclature introduced thus far, we have:

- **OrTang** – objects are oriented tangles in $B^3$, and morphisms are oriented surface cobordisms in $B^4$ up to isotopy.

- **OrPD** – objects are oriented tangles in $B^3$, with generic projection in the $z$ direction, and morphisms are formal compositions (movies) of oriented surface cobordisms, each of which has at most one "singular" moment, modulo movie moves.

- **UnAb** – objects are crossingless unoriented tangles in $B^3$, and morphisms are linear combinations of abstract unoriented cobordisms, modulo local relations.

- **DisAb** – objects are crossingless disoriented tangles in $B^3$, and morphisms are linear combinations of abstract disoriented cobordisms, modulo local relations.

- $\mathrm{Kom}(\textbf{DisAb})$ – objects are complexes in $\mathrm{Mat}(\textbf{DisAb})$, and morphisms are chain maps modulo chain homotopy.

**Lemma 2.2** *In either the oriented or unoriented context, the functor $(i \circ f)$: **PD** → **Tang**, which first forgets the ordering data on a planar diagram in **PD**, then includes the diagram into **Tang**, (recall tangles in **PD** have generic projections, whereas tangles in **Tang** need not) is a natural isomorphism of categories. (The same result is true in the disoriented context too, but we don't need that for now.)*

**Proof** First, we dispense with the ordering data on objects in **PD**: consider for a moment **PD**$^{\mathrm{unordered}}$, the same category as **PD**, but without the ordering data on crossings. The forgetful functor $f$ is an equivalence of categories; its inverse (up to natural isomorphisms) can arbitrarily specify the crossing ordering, after we've noticed that all possible orderings on a diagram are isomorphic.

Next, we construct a functor $j$ which is the inverse of the inclusion $i$ (up to natural isomorphism) of **PD**$^{\mathrm{unordered}}$ into **Tang**. For every tangle $T$, choose an isotopy $I_T$ to a general position tangle $j(T)$ (object of **PD**). For every cobordism $Y\colon T_1 \to T_2$, $I_{T_2} Y I_{T_1}^{-1}$ is a cobordism from $j(T_1)$ to $j(T_2)$. Up to "second order" isotopy, we can assume that $I_{T_2} Y T_{T_1}^{-1}$ is composed of a sequence of Reidemeister moves and Morse moves. Define $j(Y)$ to be this sequence of moves.

To show that $j(Y)$ is well-defined, we must show that choosing a different second order isotopy above changes the sequence of Reidemeister and Morse moves by movie moves. This is one of the fundamental properties of movie moves. (Note that we have different versions of movie moves for **Un**, **Or**.)

To complete the proof, it is easy to show that $\{I_T\}$ comprise an invertible natural transformation between $ij$ and the identity functor on **Tang**, and that $\{j(I_T^{-1})\}$ comprise an invertible natural transformation between $ji$ and the identity functor on **PD**. □

The cobordism categories we've described above actually split up into disjoint smaller categories, indexed by the number (and possibly orientations, when relevant) of boundary points appearing on the equator of $B^3$. These categories fit together as a canopolis (as introduced by Bar-Natan [1]), that is, a planar algebra of categories (see Jones [11]). If you're unfamiliar with planar algebras or canopolises, we've included a brief summary in Appendix A.4. The planar operations are in all cases simply given by gluing, both for objects and morphisms.

It's worth pointing out how the planar operations interact with the ordering of crossings in objects of **PD**. The internal discs of a spaghetti and meatball diagram (indexing an operation of the planar algebra) come with an ordering. When we glue together objects of **PD** inside of one of these diagrams, we simply concatenate the orderings specified inside each object.

The "matrix category" construction defining Mat(**UnAb**) and Mat(**DisAb**) has an obvious analogue for canopolises; the planar operations distribute over direct sums.

Similarly, taking complexes over a category extends to a parallel construction for taking complexes over a canopolis. In any canopolis $\mathcal{C}$, we can form a new canopolis Kom($\mathcal{C}$) whose objects are complexes in $\mathcal{C}$ and whose morphisms are chain maps (or chain maps up to homotopy). To apply a planar operation to a suitable collection of complexes in Kom($\mathcal{C}$), we take the formal tensor product of the complexes (ie form a multicomplex, sprinkle signs, and collapse), then apply the specified planar operation to each object and differential. See Appendix A.5 for more details. Notice that this planar operation on complexes in Kom($\mathcal{C}$) depends on the ordering of the internal discs through the way that signs appear when we take the tensor product of complexes, even when the original canopolis was "symmetric".

One consequence of these observations is that invariance for a local model of a movie move implies invariance for that movie move embedded in any larger tangle.

## 2.3 Disoriented Khovanov homology

Our goal is to construct a map of canopolises (that is, a functor for each category, compatible with planar operations) **OrTang** $\rightarrow$ Kom(**DisAb**). We follow closely Bar-Natan's approach, except that we replace his target category Kom(**UnAb**) with Kom(**DisAb**). We'll write $[\![T]\!]$ to denote the complex in **DisAb** associated to a tangle $T$.

It follows from Lemma 2.2 that if we want to construct a functorial invariant of **OrTang** it suffices to construct a functorial invariant of **OrPD**, and to do this it in turn suffices to:

(1) Construct a complex for each planar tangle diagram (equipped with an ordering of the crossings).

(2) Construct a map of complexes for each Reidemeister move, each Morse move and each crossing reordering map.

(3) Check that the relations coming from each oriented movie move are satisfied.

We'll do the first two steps in this subsection and verify the movie move relations in Section 3.2.

**2.3.1 The complex** The objects of **OrPD** are generated via planar algebra operations by positive and negative crossings. We define the functor on single crossings as follows:



(2-2)

In both cases, disorientation marks point to the right, relative to the overall direction of the crossing. (This is just an arbitrary convention; they could be equally well face to the left.)

Observe that a positive crossing is supported in homological heights 0 and 1, while a negative crossing is supported in heights $-1$ and $0$. We denote the grading shifts on objects simply by a multiplicative factor of some power of $q$.

Next we must define the functor on morphisms of **OrPD**. The morphisms are generated (again, via planar operations) by Reidemeister moves, Morse moves and the crossing reordering map which switches the ordering of a pair of crossings. Note that Morse moves (the cup, the saddle and the cap) are already morphisms of **DisAb**, and hence also morphisms (between one term complexes) of Kom(**DisAb**), so defining the functor on Morse moves is trivial.

When switching the ordering of a pair of crossings in a tangle, we associate a chain map which is simply $\pm 1$ on every object in the complex. Following the homological conventions described in Section A.6.2, this map is $-1$ on objects in which both crossings have been resolved in the disoriented way, and $+1$ otherwise.

In the following sections, in which we describe the chain maps associated to Reidemeister moves, we'll restrict our attention to one particular ordering of the crossings in the source and target tangle. The chain maps associated to other moves with other orderings are simply obtained by pre- and post-composition with the reordering maps from the previous paragraph.

Specifying the chain maps for the various Reidemeister moves will occupy the remainder of this subsection. Each of these chain maps will be invertible up to chain homotopy, so by the end of this subsection we will have established the following weak result: If two planar tangle diagrams are isotopic, then the complexes we assign to them are isomorphic up to chain homotopy. Full functoriality will not be established until we have verified the movie move relations in Section 3.2.

**2.3.2 The R1 chain maps** The "twist" and "untwist" chain maps for the R1a and R1b moves are shown in Figure 5 and Figure 6. The horizontal straight arrows are the differentials in the complex, and the vertical (green) arrows show the chain map itself.

Being extra careful, we might want to distinguish two variations of each of R1a and R1b, depending on whether the kink lies on the left or the right side. However, the chain maps are just mirror images of those shown here.

**2.3.3 The R2 chain maps** The Reidemeister 2 move comes in four variations, which we'll call R2al, R2ar, R2b+ and R2b−.
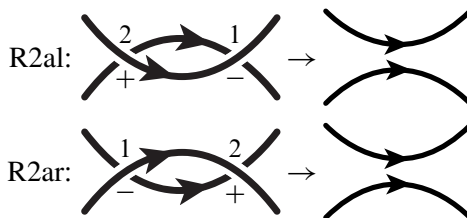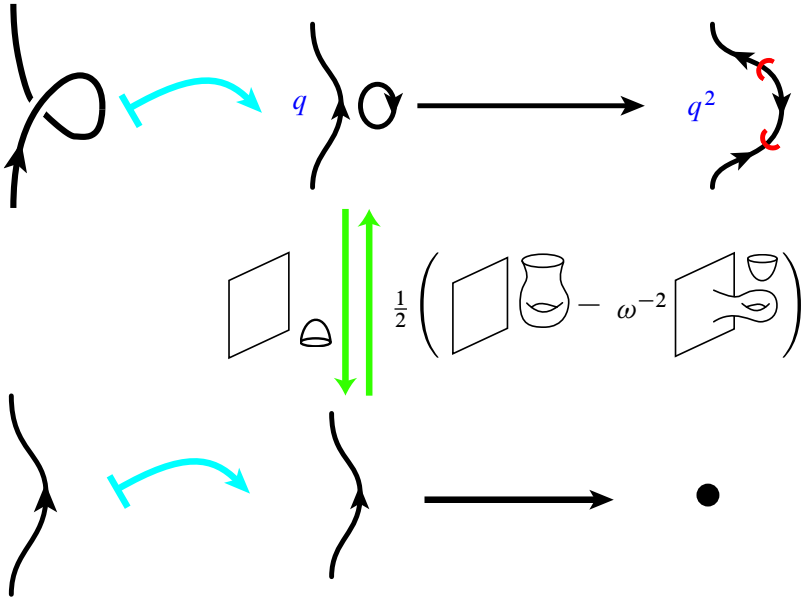
Figure 5: The R1a chain maps
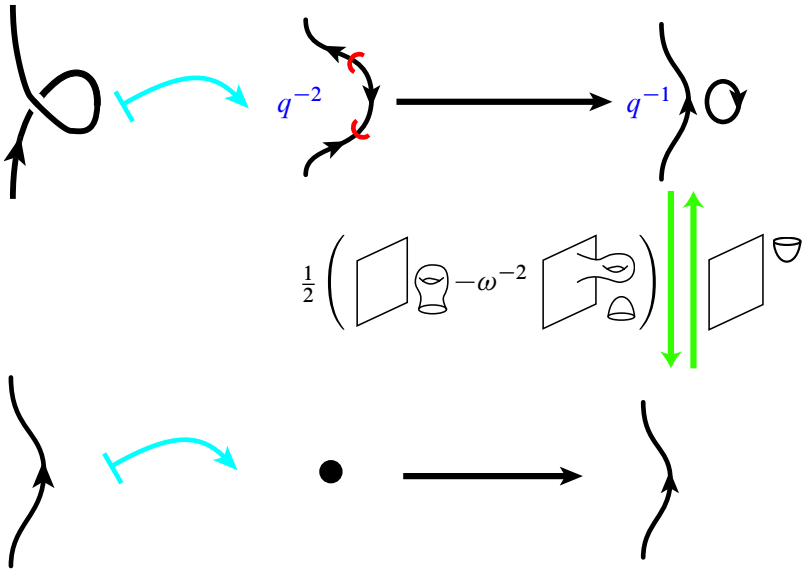


Figure 6: The R1b chain maps
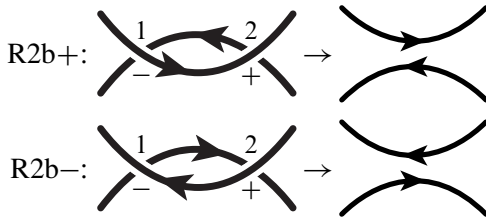
Notice that we always chose to number the crossings so the negative crossing comes first. This is, of course, an arbitrary choice, but made so that the two R2a maps, and the two R2b maps, look as similar to each other as possible.

Explicit chain maps between the two sides of the Reidemeister R2al and R2ar moves are shown in Figure 7, while maps for the R2b− and R2b+ moves are shown in Figure 8.
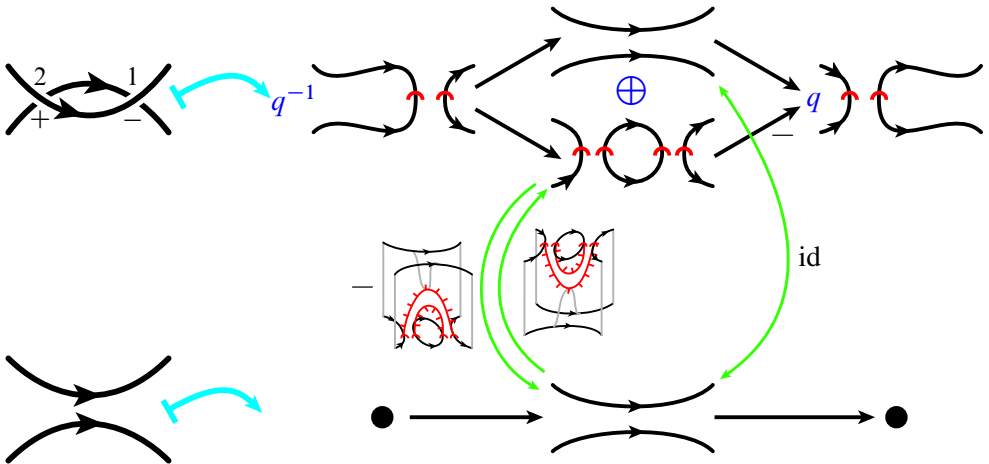


Figure 7: The R2al chain map (The R2ar chain map is identical.)

Calculations showing that these are indeed chain equivalences (and showing how to discover them in the first place) have been relegated to Appendix A.2.

**2.3.4 The R3 chain maps** The work of this section is divided into three parts. First, we explicitly describe a chain map for one variation of the R3 move, and write down several properties of this chain map. Second, we state the corresponding generalisations of these properties for the other seven variations of the R3 move. Third, we describe an alternative chain map, which is chain homotopic to the initial one, in each case.

We'll construct the chain maps for the first R3 move directly using the simplification algorithm described by Bar-Natan in [3]; specifically, applying it to the complexes
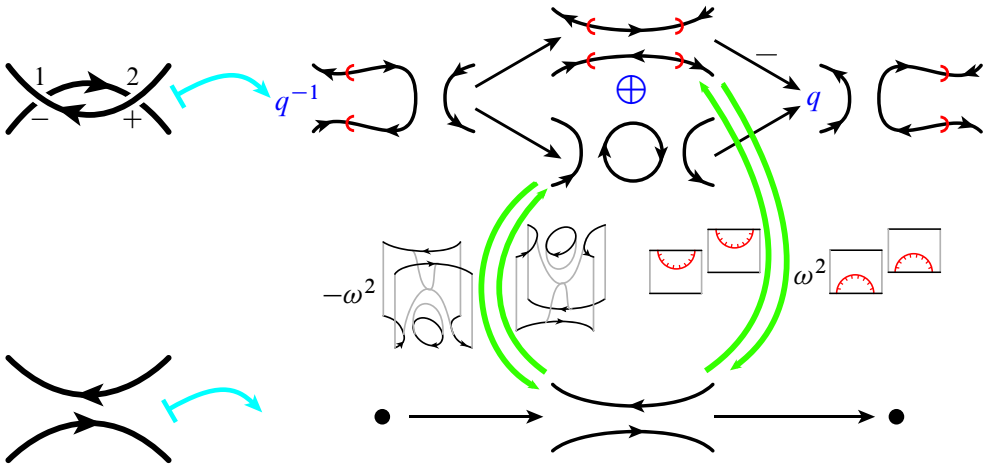
Figure 8: The R2b− chain map. (The R2b+ chain map is the same, but with all fringes reversed.)

appearing on either side of the Reidemeister move, we'll see that we obtain (almost) exactly the same complexes. Composing the "simplifying" and "unsimplifying" maps gives us the desired chain map. The result appears as Proposition 2.3.

We'll provide the chain maps for the other seven R3 moves less explicitly, using the idea that all R3 moves are equivalent modulo R2 moves.

We'll state three lemmas (Lemmas 2.4, 2.5, 2.6 for the first variation, and Lemmas 2.7, 2.8, 2.9 for the other seven variations) capturing the features of these maps relevant to later movie move calculations, but postpone the proofs until Appendix A.3.

Sadly, the "categorified Kauffman trick" first described by Bar-Natan [2] doesn't work in the disoriented category; the disorientation marks get in the way of using the second Reidemeister move. With "vertigos" (as wished for in Section 4.4), this method should recover its utility and give easier proofs of the statements we need about the seven variations, by giving an easy direct construction of the chain map in each case.

**Proposition 2.3** *There's a homotopy equivalence between the complexes associated to either side of the Reidemeister move*
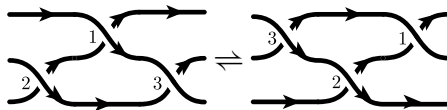
*given by*
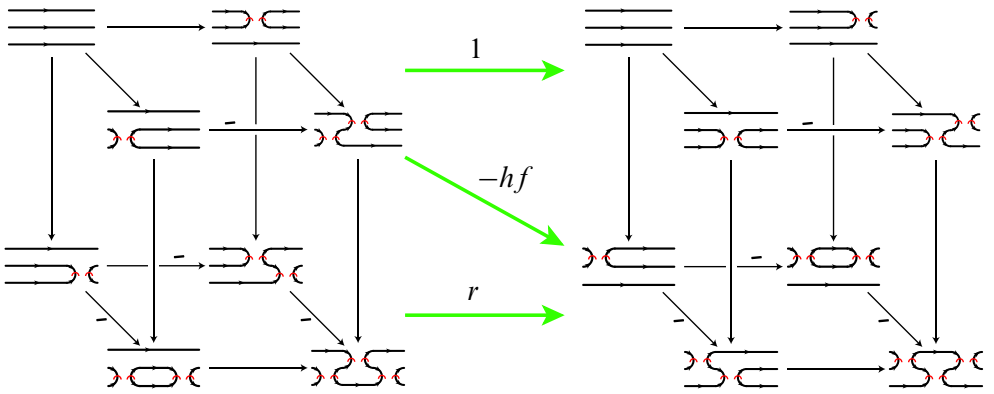


The complex for each tangle is shown as a cube, with 8 objects and 4 homological levels. The two layers, top and bottom, correspond to the two different resolutions of the highest crossing, labeled 3. The chain map providing the homotopy equivalence is the sum of the three (green) arrows each connecting one layer of the left cube to a layer of the right cube. The nonzero components of the maps are



**Remark** The names "$-hf$" and "$r$" shouldn't make any sense, unless you know about the categorified Kauffman trick, and perhaps read a future paper about the extension of Khovanov homology to disoriented tangles! If you do know the categorified Kauffman trick, we'd be considering the cones over the morphisms resolving the crossings labeled 3.

**Proof** See Appendix A.2.          □

We won't need to know much about the details of this chain map, however; what little we do is encapsulated in the following three lemmas.

**Lemma 2.4** (Needed for MM6 and 10) *The map from the bottom layer of the initial cube to the top layer of the final cube is zero.*

**Lemma 2.5** (Needed for MM6, 8 and 10) *The top layer of the initial cube is mapped identically to the top layer of the final cube.*

**Lemma 2.6** (Needed for MM6) *The leftmost and rightmost objects in the bottom layer are sent to zero. That is, the map from the bottom layer to the bottom layer kills the highest and lowest homological height pieces. Further, there is a single entry of that map, in the middle homological height, which is a multiple of the identity, that multiple is $-1$, and every other nonzero entry has a disc component attached to a circle in either the source or target object (or both).*

Now there's not just one Reidemeister 3 move; our version of Khovanov homology depends more explicitly on the orientations in the original tangle than previous constructions, and as a consequence we need to do more work. There are eight R3 moves, six "braidlike" and two "starlike". We'll name the braidlike moves by walking counterclockwise around the boundary, writing down the height of each outgoing strand. Thus in $R3_{hml}$ we see the "high" strand, the "middle" strand, then the "low strand". (We see the same sequence looking at the incoming strands.) The other braidlike moves are $R3_{hlm}$, $R3_{lhm}$, $R3_{mhl}$, $R3_{mlh}$ and $R3_{lmh}$. There are then the two starlike R3 moves, which we'll call $R3_{\circlearrowleft}$ and $R3_{\circlearrowright}$, depending on which way we have to walk around the boundary in order to see the "outgoing low", then "outgoing middle", then "outgoing high" strands. All eight Reidemeister 3 moves appear in Figure 9.
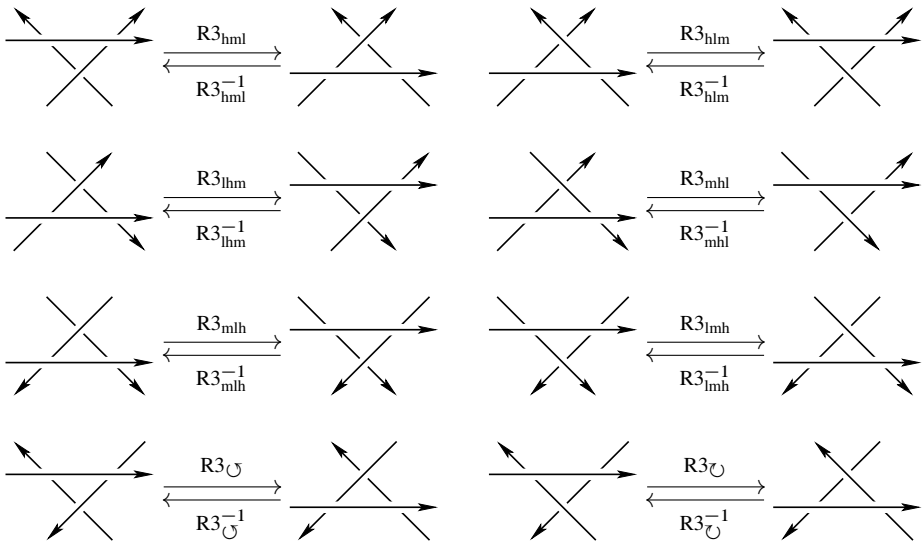
When discussing these variations of the R3 move, we'll describe the left-hand of each pair of tangles as the "initial" tangle. In every case, in the initial tangle the triangle lies to the right of the lowest strand, and in the final tangle it lies to the left. We also need to specify the ordering of the crossings in these tangles. It turns out to be convenient to use a slightly unnatural ordering: in the initial tangle we number the crossings as "middle", then "low", then "high", while in the final tangle we number them as "low", "middle", "high". Notice this rule generalises the ordering we used in describing $R3_{hml}$.
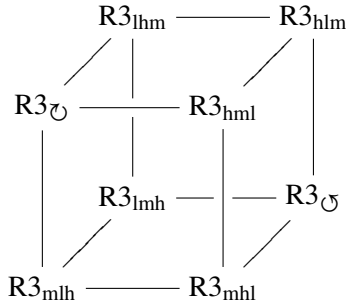


Figure 10: The cube of R3 moves

The R3 moves fit together into a cube, shown in Figure 10. The edges of this cube indicate pairs of R3 moves which are "related by R2 moves". That is, for each edge there's a commutative diagram in the category of tangles and tangle cobordisms. Here's one of the edges, connecting $R3_{\circlearrowright}$ and $R3_{hlm}^{-1}$:



(2-3)
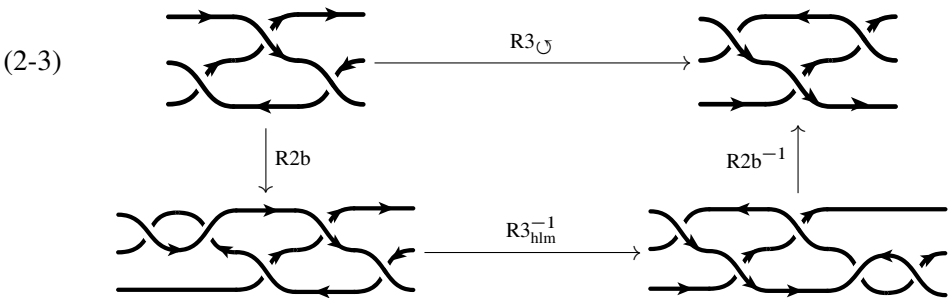
We've already specified a chain map for $R3_{hml}$, in Proposition 2.3, and we can now specify chain maps for each of the others. To do this, we pick some spanning tree for the cube. We'll now inductively define the chain map for a Reidemeister 3 variation in terms of the already defined chain map for another variation adjacent in the spanning tree. We simply write down the composition of the other three chain maps appearing in the commutative square corresponding to Equation (2-3) for the appropriate edge.

We're never going to explicitly write down all the R3 maps; it would be incredibly tedious. Instead, we'll just write down some lemmas (Lemmas 2.8, 2.7 and 2.9, generalising Lemmas 2.5, 2.4 and 2.6 respectively), which encapsulate the facts we need for the movie move calculations. We'll prove these statements by showing how they "propagate" along the edges of the cube in Figure 10.

Finally, you might worry about the choice of spanning tree. However, the sequence of movie moves corresponding to a face of the cube is a cobordism isotopic to the identity, so functoriality will eventually assure us the choice didn't matter.

In order to state our more general lemmas, we'll need to describe various parts of the complexes appearing on either side of the variations of the R3 moves. Thinking of such a complex as a cube, as in Proposition 2.3, we'll consider it as split into two layers, corresponding to the two resolutions of the "highest" crossing (that is, the crossing between the "high" and "middle" strands). While we could describe the two layers as the "oriented" layer and the "disoriented layer", there's something more useful; we'll describe them as the "orthogonal" ($\mathcal{O}$) and "parallel" ($\mathcal{P}$) layers, as shown in Figure 11, depending on whether the strands in the resolution of the highest crossing are orthogonal or parallel to the third strand not involved at the resolved crossing.

$$\underset{\times}{\nearrow} = C\left(\left(\mathcal{O} = \underset{\times}{\nearrow}\right) \overset{s}{\to} \left(\mathcal{P} = \underset{\times}{\nearrow}\right)\right)$$

Figure 11: The complex for each tangle appearing in an R3 variation can be divided into two layers, the "orthogonal" ($\mathcal{O}$) and "parallel" ($\mathcal{P}$) layers.

Notice that the differentials in the cube between the two layers point either from the $\mathcal{O}$ layer to the $\mathcal{P}$ layer, or from the $\mathcal{P}$ layer to the $\mathcal{O}$ layer. This depends on which R3 variations we're looking at, in particular on the sign of the highest crossing, and whether its oriented resolution is orthogonal or parallel to the third strand. Notice that order of layers alternates between $\mathcal{O} \to \mathcal{P}$ and $\mathcal{P} \to \mathcal{O}$ as we step across any edge in the cube in Figure 10.

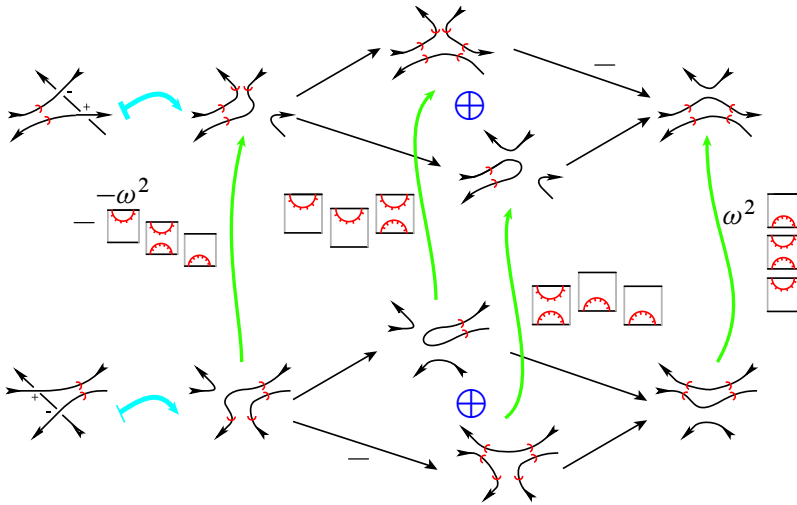We can then write each chain map R3$_\star$ (where $\star$ is one of hml, hlm, lhm, mhl, mlh, lmh, $\circlearrowleft$ or $\circlearrowright$) as the sum of four components, R3$_\star$ = R3$_\star^{\mathcal{O}\to\mathcal{O}}$ + R3$_\star^{\mathcal{O}\to\mathcal{P}}$ + R3$_\star^{\mathcal{P}\to\mathcal{O}}$ + R3$_\star^{\mathcal{P}\to\mathcal{P}}$, where R3$_\star^{a\to b}$ are the components from the $a$ layer to the $b$ layer.

**Lemma 2.7** *If the layers of R3$_\star$ are arranged as $\mathcal{O} \to \mathcal{P}$, then the map from the parallel layer to the orthogonal layer, R3$_\star^{\mathcal{P}\to\mathcal{O}}$, is zero. Otherwise, if the layers are arranged as $\mathcal{P} \to \mathcal{O}$, then the map R3$_\star^{\mathcal{O}\to\mathcal{P}}$ is zero. (That is, the diagonal map pointing backwards in homological height is always zero.)*

**Lemma 2.8** *The map between the orthogonal layers, $R3_\star^{\mathcal{O}\to\mathcal{O}}$, is the identity chain map, when $\star = $ hml, lhm, mhl or lmh. When $\star = $ hlm, mlh, $\circlearrowright$ or $\circlearrowleft$, the maps $R3_\star^{\mathcal{O}\to\mathcal{O}}$ are nonzero multiples of a certain standard chain map; forgetting disorientation data and coefficients, this map is the identity chain map. The disorientation seams are the minimal ones compatible with the boundary disorientation marks. The coefficients are all either $-1$ or $\omega^2$, and are determined by the rule that the coefficient $\kappa_\star$ of the chain map in the lowest homological height is given by*

$$\kappa_\star = \begin{cases} \omega^2 & \text{if } \star = \text{hlm or } \circlearrowleft, \\ -1 & \text{if } \star = \text{mlh or } \circlearrowright. \end{cases}$$

*The fact that these are chain maps then determines the other coefficients; in particular, on the highest homological height the coefficient is $-\omega^2\kappa_\star$.*

**Remark** This dichotomy distinguishes whether the orthogonal layer of the cube comes from an oriented or disoriented resolution of the highest crossing. These data are displayed in the last column of Table 1.

As an example, the map $R3_{\text{hlm}}^{\mathcal{O}\to\mathcal{O}}$ is:



Notice here that the inverse map is obtained by taking the adjoint (reflection in the time direction) of each disoriented surface, and moving the coefficient of $\omega^2$, but not the coefficient of $-1$, to the other component in homological height 0.

As another example, the map $R3_{\circlearrowleft}^{\mathcal{O}\to\mathcal{O}}$ is:



Again, the inverse map is obtained by taking the adjoint, and moving the coefficient of $\omega^2$ (but not the coefficient of $-1$), appearing at height $0$ over to the other map at that height.

**Lemma 2.9** *The maps between the parallel layers, $R3_{\star}^{\mathcal{P}\to\mathcal{P}}$, kill the highest and lowest homological heights. Further, in the middle homological height there are a pair of objects (one in the source complex, one in the target complex) which have the same unoriented diagram, and the component of the $R3_{\star}^{\mathcal{P}\to\mathcal{P}}$ map between these is the unique disoriented surface with minimal disorientation seams, and a coefficient of*

$$
p_{\star} = \begin{cases}
-1 & \text{if } \star = \text{hml or lmh}, \\
1 & \text{if } \star = \text{hlm or mlh}, \\
\omega^2 & \text{if } \star = \text{lhm or mhl}, \\
-\omega^2 & \text{if } \star = \circlearrowright \text{ or } \circlearrowleft.
\end{cases}
$$

*Every other entry of the map in the middle homological height is some multiple of a surface with a disc component attached to a circle in either the source or target object (or both).*

The proofs appear in Appendix A.3.

At this point we can also give a description of the inverses of these chain maps.

**Corollary 2.10** *Lemmas 2.7, 2.8 and 2.9 also hold without changes when describing the inverses of the R3 chain maps.*

| R3 variation | highest crossing | order of layers | orthogonal layer is |
|---|---|---|---|
| $R3_{hml}$ | $+$ | $\mathcal{O} \to \mathcal{P}$ | oriented |
| $R3_{hlm}$ | $+$ | $\mathcal{P} \to \mathcal{O}$ | disoriented |
| $R3_{lhm}$ | $+$ | $\mathcal{O} \to \mathcal{P}$ | oriented |
| $R3_{mhl}$ | $-$ | $\mathcal{P} \to \mathcal{O}$ | oriented |
| $R3_{mlh}$ | $-$ | $\mathcal{O} \to \mathcal{P}$ | disoriented |
| $R3_{lmh}$ | $-$ | $\mathcal{P} \to \mathcal{O}$ | oriented |
| $R3_{\circlearrowleft}$ | $-$ | $\mathcal{O} \to \mathcal{P}$ | disoriented |
| $R3_{\circlearrowright}$ | $+$ | $\mathcal{P} \to \mathcal{O}$ | disoriented |

Table 1: The variations of the R3 move

**Proof** Consider the operation of rotating a tangle by $\pi$, and reversing all orientations. Notice that this interchanges the source and target tangles of each R3 variation.

Being a little more careful, and thinking about the source and target tangles with their specified ordering of crossings, this operation actually needs to be followed by switching the ordering of the low and middle crossings.

Thus for each R3 variation, we produce a chain map pointing the opposite direction, by rotating each component of the original chain map by $\pi$, reversing all orientations and disorientations, and introducing an extra sign in each component going between a pair of resolutions in which for one or the other of the initial and final resolutions, but not both, both the low and middle crossings have been resolved in the disoriented way.

We now make two claims. Firstly, that this chain map really is the inverse of the original map, and secondly, that this chain map is correctly described by Lemmas 2.7, 2.8 and 2.9.

First, we consider the $\mathcal{O} \to \mathcal{O}$ parts of the map. It is readily seen (trivial in the cases hml, lhm, mhl or lmh, easy in the cases hlm and mlh, and requiring an easy calculation involving disorientations in the cases $\circlearrowright$ and $\circlearrowleft$) that at the lowest homological height, the composition of the original map and the candidate inverse is the identity. This is enough to know that the candidate really is the inverse.

Second, Lemma 2.7 holds obviously, Lemma 2.8 holds because the signs introduced by reordering occur at homological height 0, so cannot affect the sign $\kappa_\star$, and Lemma 2.9 holds because the reordering signs occur at heights $\pm 1$, so cannot affect the sign $p_\star$.

Notice that the inverses of the example $\mathcal{O} \to \mathcal{O}$ maps given above agree with the description here. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

The third task of this section is to describe an alternative chain map for each Reidemeister 3 move. This alternative will be chain homotopic to the one described above, but not identical.

The mirror image (in the direction perpendicular to the plane) of a tangle is simply the obvious topological operation. At the level of the corresponding Khovanov complexes, this corresponds to negating the homological height of each step of the complex, and replacing each differential with its time reverse, by switching source and target. That is, the mirror image of a complex $(C^\bullet, d)$ is $(\bar{C}^\bullet, \bar{d})$, with $\bar{C}^i = C^{-i}$, and $(\bar{d}_i \colon \bar{C}^i \to \bar{C}^{i+1}) = (d_{-i-1} \colon C^{-i-1} \to C^{-i})^*$, where the $*$ here means time reversal, or "adjoint". By the mirror image of a chain map $f^\bullet$, we mean $\bar{f}^\bullet$, with $\bar{f}^i = f^{-i}$; that is, exactly the same components, but each in negated homological height.

We can think of the alternative chain map in two different ways. First, and secretly, we think of it as coming from performing the Kauffman trick on the lowest crossing, rather than the highest crossing as above. Second, we can simply think of it, and define it, as the mirror image of one of the chain maps above. Actually, more precisely, we need to modify this mirror image in two ways. First, in all cases, we must pre- and post-compose with crossing reordering maps, to ensure that we start and finish at the same ordered tangles as the usual chain maps. Second, only for the starlike R3 variations, we need to multiply the mirror image chain map by $-\omega^2$. (This will ensure that the mirror image chain map really is homotopic to the usual one. Recall of course that in the disoriented theory, $-\omega^2 = 1$!)

Notice that taking mirror image exchanges pairs of R3 move variations, switching the labels "h" and "l", and interchanging ↺ and ↻. Thus $R3_{hml}$ and $R3_{lmh}$, which are antipodal in the cube of R3 variations in Figure 10, are exchanged, as are $R3_↺$ and $R3_↻$. The other pairs are $R3_{lhm}$ and $R3_{hlm}^{-1}$, and $R3_{mlh}$ and $R3_{mhl}^{-1}$, which are each adjacent in the cube.

To distinguish the chain maps defined in this way from the ones described above, we'll write a bar over the top. Thus $\overline{R3}_{hml}$ is defined by taking the chain map for $R3_{lmh}$, and applying the mirror image operation described in the paragraph above, and reordering crossings in the source and target tangles appropriately.

Passing to the mirror image move reverses the "order of layers" appearing in Table 1. It's easy to see that the mirror image of a chain map for one vertex does not give the chain map for the opposite vertex which has been described above. This is essentially because the lemmas above are written in terms of the orthogonal and parallel layers with respect to the highest crossing, which are not preserved by mirror image. For example, look at the pair $R3_{hml}$ and $R3_{lmh}$, and in particular the completely oriented resolution. The completely oriented resolution is in the orthogonal layer for $R3_{hml}$, so

the chain map above acts as the identity here, by Lemma 2.7 (or indeed, the original special case Lemma 2.4). However the completely oriented resolution is killed by the usual chain map for $R3_{lmh}$, being in the parallel layer, using Lemma 2.9. Thus we see that the chain map for $R3_{hml}$ coming from the mirror image of the chain map for $R3_{lmh}$ is in fact different from the usual one.

On the other hand, these maps turn out to be homotopic to the usual maps, even though we have seen they are not equal on the nose. The argument relies on two results which live more naturally later in the paper, namely Corollary 3.3, appearing in the next section, and Lemma 3.5 appearing in Section 3.2, so the reader may prefer to postpone deciphering this argument until having reached those statements! Corollary 3.3, appearing in the next section, assures us that the relevant space of chain maps, up to homotopy, is 1 dimensional. Thus we know that each mirror image map must be homotopic to some multiple of the usual map, and we only need to show that multiple is always 1. To do this, we look at a particular resolution, namely the unique resolution which is in an extreme homological height of the $\mathcal{O}$ layer for both the usual map and the mirror image map. Lemma 2.8 then describes how this resolution is mapped to the corresponding resolution of the target tangle, and it suffices, by Lemma 3.5 to check that both the usual map and the mirror map act in the same way, without coefficients, on this resolution. That check follows directly from Lemma 2.8, along with the relevant crossing reordering calculations. Recall also the coefficient of $-\omega^2$ which we smuggled into the definition of the mirror image maps for the starlike moves, precisely to allow the present result.

An important point we need to make is that the three Lemmas 2.7, 2.8 and 2.9 still apply to the mirror image maps, replacing as needed each reference to an R3 variation $R3_\star$ with $\overline{R3_{\star'}}$, where $\star'$ is the mirror variation, and understanding "orthogonal" and "parallel" layers as referring to the layers given by resolving the lowest, rather than the highest, crossing.

In particular, in regard to Lemma 2.8, while $R3_\star^{\mathcal{O} \to \mathcal{O}}$ is the identity chain map when $\star = \mathrm{hml}, \mathrm{lhm}, \mathrm{mhl}$ or $\mathrm{lmh}$, when we look at $\overline{R3_\star^{\mathcal{O} \to \mathcal{O}}}$, it is $\star = \mathrm{hml}, \mathrm{lmh}, \mathrm{mlh}$ and $\mathrm{hlm}$ that give the identity. This will be important in the discussion of movie move 6.

# 3 Checking movie moves

## 3.1 Duality and dimensions of spaces of chain maps

Most nice (or at least, interesting to topologists) monoidal categories have duals. There are many formulations of this; see for example Barrett and Westbury [4] for "pivotal

categories", etc. The category $\mathcal{C}$ should have an involution $*$ on objects, called the dual, and isomorphisms between hom-sets of the form

$$\mathrm{Hom}_{\mathcal{C}}\left(U \otimes V, W\right) \cong \mathrm{Hom}_{\mathcal{C}}\left(U, W \otimes V^*\right)$$

(along with the three other obvious variations of this), satisfying some axioms (corresponding diagrammatically to "straightening an S-bend").

There's no shortage of examples. Categories of diagrams up to isotopy (see Joyal and Street [12]) are generally tautologically equipped with duals, given by $\pi$ rotations, and the natural isomorphisms between hom-sets are just planar isotopies. Categories of representations of quantum groups have duals, provided by the antipode in the Hopf algebra structure of the quantum group. Bimodules over a von Neumann algebra have duals; there the isomorphism between hom-sets is called "Frobenius reciprocity"; see Bisch [5].

We'll prove a result along these lines here. To fit with the above pattern, briefly consider the 2-category whose objects are (oriented) points on a line, whose 1–morphisms are tangles between these points, and whose 2–morphisms are chain maps up to homotopy between the Khovanov complexes associated to the tangles. There's a duality functor, at least at the level of 0- and 1–morphisms, given by reflection. We'll prove that there are isomorphisms of the type described above.

In our case there is more structure than in the above examples, since we're actually in a 3– or 4–category rather than a 2-category. (3-category if we're thinking in terms of tangle projections living in $B^2$; 4–category if we're thinking in terms of unprojected tangles living in $B^3$, with cobordisms in $B^4$.) More specifically, suppose we have tangles $P$ with $l+m$ points on its boundary and $Q$ with $m+n$ points on its boundary. We can glue $P$ and $Q$ together along $m$ consecutive points to obtain a new tangle, denoted $P \bullet Q$, with $l+n$ points on its boundary.

**Proposition 3.1** *Given oriented tangles $P$, $Q$ and $R$, with $l+m$, $m+n$ and $l+n$ points, respectively, on their boundaries, there is an isomorphism between the spaces of chain maps up to homotopy*

$$F: \mathrm{Hom}_{\mathrm{Kh}}\left(\llbracket P \bullet Q \rrbracket, \llbracket R \rrbracket\right) \cong \mathrm{Hom}_{\mathrm{Kh}}\left(\llbracket P \rrbracket, \llbracket R \bullet \bar{Q} \rrbracket\right)\left\{\frac{m-n}{2}\right\}.$$

*($\bar{Q}$ denotes the reflection of $Q$.)*

*Diagrammatically, this statement claims that there's an isomorphism between the spaces of chain maps we can fill inside the following two cylinders.*
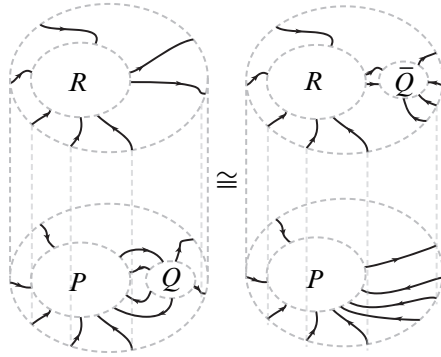


*These isomorphisms are natural in the sense that they are compatible with pre-composition with a morphism into $P$, and with post-composition with a morphism out of $R$.*

**Remark**  For now, we're just claiming that there is some isomorphism; in particular, all we'll need for now is that the dimensions of the morphisms spaces are the same.

In a future paper, we'll explain a coherence result for these isomorphisms. Essentially this result is the difference between "functoriality in $B^3$" and "functoriality in $S^3$". There are pairs of cobordisms in $B^3$ which are not isotopic in $B^3$, but become isotopic in $S^3$. The coherence result for the maps described in the proposition above requires us to show that such pairs give homotopic chain maps, and this remains beyond the scope of the current paper.

**Proof**  We'll prove the result for a short list of (very!) small tangles $Q$, which easily imply the rest. Namely $Q = \times, \times, )$ and $($, and the other oriented versions of these tangles. We can then build the isomorphism for an arbitrary $Q$ by composing isomorphisms for the constituent pieces of the tangle $Q$.

We'll begin with $Q = \times$, a negative crossing oriented to the right. (The case for a positive crossing is exactly analogous.) Given a chain map

$$f \in \mathrm{Hom}_{\mathrm{Kh}}\left(\left[\!\!\left[P \bullet \times\right]\!\!\right], [\![R]\!]\right),$$
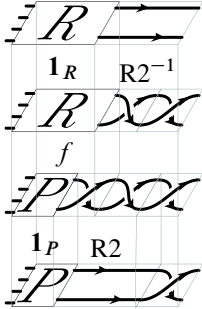
we'll produce the chain map

$$F(f) = (f \bullet \mathbf{1}_{\times}) \circ (\mathbf{1}_P \bullet \mathrm{R2}) \in \mathrm{Hom}_{\mathrm{Kh}}\left([\![P]\!], \left[\!\!\left[R \bullet \times\right]\!\!\right]\right).$$

We propose that the inverse of this construction is given by
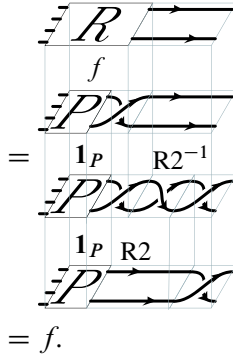
$$\mathrm{Hom}_{\mathrm{Kh}}\left(\llbracket P \rrbracket, \left\llbracket R \bullet \asymp \right\rrbracket\right) \ni g \mapsto F^{-1}(g) = (\mathbf{1}_R \bullet \mathrm{R2}^{-1}) \circ (g \bullet \mathbf{1}_{\asymp}).$$

The composition $F^{-1} \circ F$ applied to a chain map $f$ is:

$$(\mathbf{1}_R \bullet \mathrm{R2}^{-1}) \circ (((f \bullet \mathbf{1}_{\asymp}) \circ (\mathbf{1}_P \bullet \mathrm{R2})) \bullet \mathbf{1}_{\asymp}) =$$



To see that this just $f$, we can do some tensor category arithmetic:

$$F^{-1}(F(f)) = (f \bullet \mathbf{1}_{\longrightarrow}) \circ (\mathbf{1}_P \bullet (\mathbf{1}_{\asymp} \bullet \mathrm{R2}^{-1} \circ \mathrm{R2} \bullet \mathbf{1}_{\asymp}))$$

$$=$$



$$= f.$$

The critical step in this calculation came at the end, in claiming that

$$(\mathbf{1}_{\asymp} \bullet \mathrm{R2}^{-1}) \circ (\mathrm{R2} \bullet \mathbf{1}_{\asymp}) = \mathbf{1}_{\asymp}.$$

This is exactly checking MM9, the ninth movie move. Although it strains the logical order of the paper somewhat, we'll postpone that calculation until Section 3.2.2, where we do all the other movie moves, being careful to point out that we don't use any of the results of this section while checking MM9.

A very similar argument shows $F(F^{-1}(g))$ is also just $g$.

The case $Q = \asymp$ is very similar.

Next, we deal with the case that the tangle $Q$ is just an arc, $\subset$. This time, the map $F$ is given by

$$F(f) = (f \bullet \mathbf{1}_{\supset}) \circ (\mathbf{1}_P \bullet \bigcup),$$

with inverse

$$F^{-1}(g) = (\mathbf{1}_R \bullet \text{⟧}) \circ (g \bullet \mathbf{1}_{\subset}).$$

The argument that $F$ and $F^{-1}$ are inverses is even easier than before; some formal tensor category arithmetic and cobordism arithmetic is all we need. For example,

$$F(F^{-1}(g)) = \boxed{\begin{array}{c} P \\ \mathbf{1}_P \\ P \\ g \\ R \\ \mathbf{1}_R \\ R \end{array}} = g.$$

The other three cases where $Q$ is an arc are very similar. □

We now get an easy corollary, which you should think of as a nice analogue of Bar-Natan's result about simple tangles in [2].

**Corollary 3.2** *Let $T_1$ and $T_2$ be tangles with $2k$ endpoints such that $\overline{T_1}T_2$ is an unlink with $m$ components. Then the space of chain maps modulo chain homotopy from $[\![T_1]\!]$ to $[\![T_2]\!]$ in grading $m-k$ is 1–dimensional, and all chain maps of grading higher than $m-k$ are chain homotopic to zero.*

**Proof** By Proposition 3.1

$$\begin{aligned} \mathrm{Hom}_{\mathrm{Kh}}(T_1, T_2) &\cong \mathrm{Hom}_{\mathrm{Kh}}(\varnothing, \overline{T_1}T_2)\{-k\} \\ &\cong [\![\overline{T_1}T_2]\!]\{-k\} \\ &\cong (\mathcal{R}\{-1\} \oplus \mathcal{R}\{+1\})^{\otimes m}\{-k\} \end{aligned}$$ □

The next corollary is well known in the field, but perhaps worth stating again.

**Corollary 3.3** *The chain maps defined for the three Reidemeister moves in Section 2.3 are, up to chain homotopy and scalar multiples, the unique chain maps between the complexes in the appropriate grading.*

## 3.2 Movie moves

In this section, we'll complete the proofs of Theorem 1.1 and Theorem 1.2, by checking that changing the presentation of a cobordism by a movie move does not change the associated chain map.

We'll first prove some preparatory lemmas, which will significantly reduce the computational burden.

**Definition 3.4** Let $C^\bullet$ and $D^\bullet$ be complexes in some additive category, with $A$ a direct summand in some $C^i$. We say $A$ is $C - D$ *homotopically isolated* if, for any homotopy $h: C^\bullet \to D^{\bullet-1}$, the restriction of $dh + hd$ to $A$ is zero.

If we're in a graded category then $A$ is $C - D$ homotopically isolated if $dh + hd$ is zero for every grading $0$ homotopy $h$.

**Lemma 3.5** *Say $C^\bullet$ and $D^\bullet$ are complexes associated to two tangle diagrams (thus, complexes in the category of abstract disoriented cobordisms), and say $A$ is a smoothing appearing as a direct summand of some step of the complex $C^\bullet$. Then*

(1) *$A$ is $C - C$ homotopically isolated if $A$ does not contain any loops, and is not connected by differentials to any diagrams containing loops;*

(2) *$A$ is $C - D$ homotopically isolated if $C^\bullet$ and $D^\bullet$ do not contain any diagrams with loops.*

**Proof** This is easy from our definition of gradings on morphisms. A homotopy $h: B \to A$ always maps backward one step in homological height (regardless of whether $A$ and $B$ are in the same or different complexes). Thus $h$ will have "bare" grading $+1$, but there are no positive grading morphisms between loopless diagrams by Euler characteristic considerations. □

**Lemma 3.6** *In each of movie movies 6 through 8, every smoothing of the complex $C$ associated to the initial frame is $C - C$ homotopically isolated. In movie moves 11, 13 and 15, every smoothing in the initial and final complexes, $C$ and $D$, is $C - D$ homotopically isolated.*

**Proof** This is trivial; no loops occur anywhere in these complexes. □

We don't need to say anything about homotopy isolation in MM9, because we won't be using any of these simplifying lemmas in that case—instead, the complete calculations are necessary for the sake of Proposition 3.1 on duality for Khovanov homology.

We can't say anything about homotopy isolation in MM12 and MM14, because, when reading backwards in time, there aren't any isolated objects! We will also use homotopy isolation in MM10, but identifying a different smoothing in each of the many variations; the details are in Section 3.2.2.

**Lemma 3.7** *Suppose $f$ and $g$ are chain maps between the complexes $[\![T_1]\!]$ and $[\![T_2]\!]$, and we know $f \simeq \alpha g$ for some $\alpha \in \mathbb{Z}[\frac{1}{2}, \omega]$. If $f$ and $g$ agree and are nonzero on some homotopically isolated object in the complex $[\![T_1]\!]$, say $O$, then in fact $f \simeq g$ are actually homotopic.*

**Proof** On $O$, $f - \alpha g = dh + hd = 0$, so $f = \alpha g = g$. Thus $\alpha$ must be 1, so $f$ and $g$ are homotopic.                                                                          □

Finally, we observe that Corollary 3.2 applies to every movie move. The join of the initial and final tangle is always just an unlink, so the relevant space of chain maps modulo homotopy is always one dimensional. Combined with the lemmas above, we see that every movie move must come out right up to a multiple (in $\mathbb{Z}[\frac{1}{2}, \omega]$), and to detect this multiple we can simply look at the restriction of the map to a single homotopically isolated object. (Remembering, of course, that MM9, MM12, and MM14 take a little more work; MM9 because there we don't have access to any of the results on duality, in particular Corollary 3.2, and MM12 and MM14 because we can't find homotopically isolated objects in the reverse time direction.) Movie moves MM1 through MM10 describe isotopies, not general cobordisms, so there any multiple would actually have to be a unit.

In the calculations for MM6, MM8, MM9 and MM14, we'll explicitly keep track of the ordering of the crossings. In all of the other calculations, it turns out the ordering of crossings is irrelevant; using the tricks described above, we only need to look at the action of the chain maps on part of the complex, and in most cases any crossing reordering maps automatically act on the objects we're interested in by $+1$, simply because there's at most one crossing which has been resolved disorientedly.

**3.2.1 MM1–5** The first five movie moves are trivial; they simply say that a Reidemeister move followed by its inverse is the identity.

**3.2.2 MM6–10** Movie moves 6 through 10 involve no Morse moves, and so are reversible. We only need to check one time direction.

In the following calculations (and those for MM11–15), red and purple bands appearing in diagrams in complexes are simply a hint to the reader, marking where crossings appeared in the original tangle. (We hope they don't obscure too much for reader looking at a black and white printout.)

**MM6**
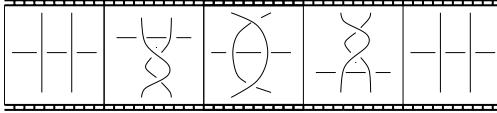


There are 24 variations of MM6. To see this we'll first of all make use of rotational symmetry to require that the "horizontal" strand (the one not involved in either R2 move) points from left to right. There are then sixteen possibilities for the initial frame of the movie move; these come from four choices of height orderings and four choices of orientations. The horizontal strand can either lie entirely above or entirely below the two vertical strands ("noninterleaved"), or it may pass under one and over the other ("interleaved", "ascending" or "descending"). The two vertical strands may be either parallel or antiparallel. When they are parallel, they may point up or down, and when they are antiparallel they may have a clockwise or anticlockwise orientation. All of these variations are displayed in Figure 12.
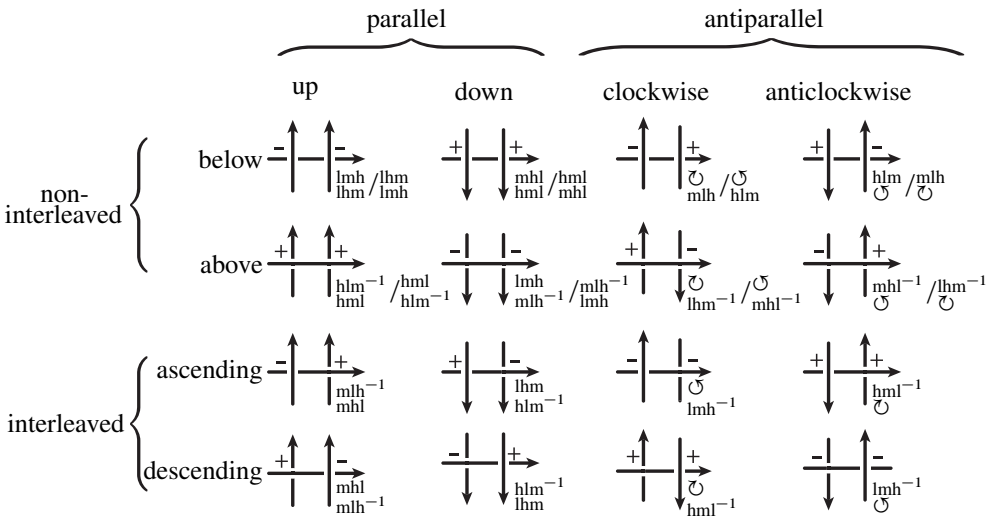


Figure 12: 16 variations for the initial frame of MM6

Note that the interleaved variations were not treated at all in version 1 of this paper on the arXiv, or in Caprau's paper [6] on the disoriented version of Khovanov homology.

Further, the eight variations in which the strands are "noninterleaved" (the first two rows of Figure 12) each have two subvariations, which we don't see until the second frame of the movie Of the two vertical strands, either one can pass above the other during the R2 moves; in Figure 12, the "left passing above the right" subvariation is listed to the left of the slash. In the "interleaved" variations, there is no choice here.

We will thus treat four major cases:

- noninterleaved, parallel variations,
- noninterleaved, antiparallel variations,
- interleaved, parallel variations and
- interleaved, antiparallel variations.

Noninterleaved parallel variations: There are four possible initial frames which are "noninterleaved" and have parallel vertical strands. Each of these initial frames has two possible subvariations, depending on the relative heights of the vertical strands during the R2 moves. For each of the four initial frames, we will treat uniformly the subvariations in which the upper R2–induced crossing is negative and the lower one is positive, and then indicate how to treat the other four subvariations.

Recall that our lemmas encapsulating the details of the R3 variations require that we separate the initial and final complexes into layers $\mathcal{O}$ and $\mathcal{P}$ by resolving a crossing. Maneuvering through the pair of R3s in this movie move is most efficiently managed by resolving the R2–induced crossings: the upper one for the first R3, and the lower one for the second R3. Notice that since the upper crossing is negative, the first R3 will have homological ordering $\mathcal{O} \to \mathcal{P}$, while the second R3 will have ordering $\mathcal{P} \to \mathcal{O}$. It's also worth mentioning that the horizontal strand could be above or below the vertical ones, meaning that these two crossing could be either the high or low crossings in their respective R3 moves. However, Lemmas 2.7, 2.8, and 2.9 work regardless[2] of whether the resolved crossing is high or low, so we needn't treat them separately.

Our "bundle" of maps for this subcase, then, will look like this:

(3-1)
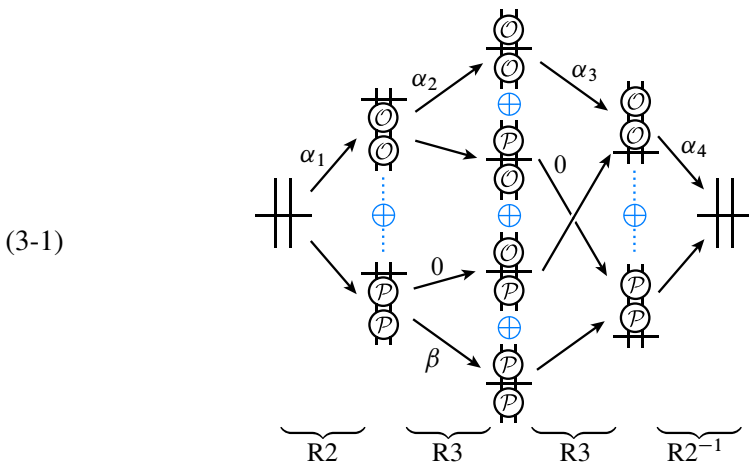
In this diagram, the $\mathcal{O}$s and $\mathcal{P}$s describe whether the indicated crossing resolution has strands orthogonal or parallel to the horizontal strand. For example



is our notation for
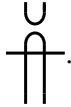


Also, we've cheated slightly with this diagram: the fourth column should contain two additional summands, those with mixed $\mathcal{O}$s and $\mathcal{P}$s. However, while there are nonzero maps into these summands, the $\mathrm{R2}^{-1}$ maps out are always zero. Thus we needn't excessively complicate things with their presence.

We're left with a sum of four compositions. The two middle compositions are both zero, as each contains a leg (labelled with "0") that's zero by Lemma 2.7. The top composition ($\alpha_i$'s) is just the identity: $\alpha_1$ and $\alpha_4$ are components of R2a moves, and $\alpha_2$ and $\alpha_3$ are each the identity, by Lemma 2.8 (each map is a component of the $\mathcal{O} \to \mathcal{O}$ map; when the horizontal strand lies below, the R3 moves are lmh lhm, mhl and hml, which are exactly the four for which the $\mathcal{O} \to \mathcal{O}$ part of the R3 map is the identity, and when the horizontal strand lies above, the R3 moves are hml, hlm, lmh and mlh, which are exactly the four for which the $\mathcal{O} \to \mathcal{O}$ part of the mirror image R3 map is the identity). The bottom composition is slightly more mysterious, but we see that the map $\beta$ sends homologically extreme smoothings to zero by Lemma 2.9. Thus, if we choose an extreme smoothing to begin with, for example the doubly oriented one, it will necessarily map to an extreme smoothing in
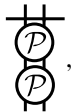


and thence to zero. Further, as mentioned before, any initial smoothing here is homotopically isolated, so the computation with this particular smoothing suffices.

The other four subvariations, in which the signs of the R2–induced crossings are reversed, are proven analogously: note that Equation (3-1) will then have all $\mathcal{O}$s and $\mathcal{P}$s swapped.

 Noninterleaved antiparallel variations: Let's consider first those cases in which the vertical strands are oriented in the anticlockwise direction, so the left vertical strand is

oriented downward, and the right upward. Again we'll be referring to Equation (3-1). Consider the smoothing
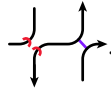


Since we are looking at noninterleaved antiparallel variations, the two signs of the initial crossings differ, and so this resolution has homologically extreme height. In particular, when the horizontal strand is below the vertical strands, this resolution has height $+1$, and when the horizontal strand is above the vertical strands, it has height $-1$.

The composition $\alpha_4 \circ \alpha_3 \circ \alpha_2 \circ \alpha_1$ then looks like
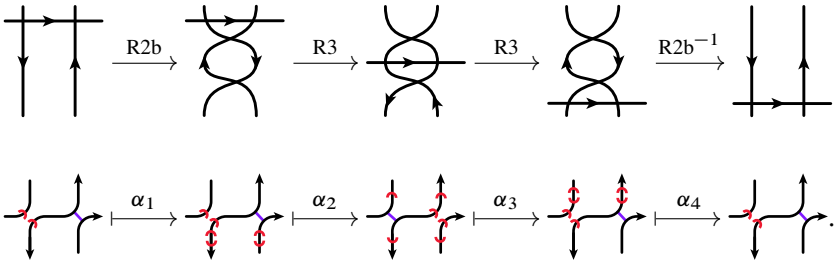


We now need to describe the maps $\alpha_i$, using our definitions of the R2 chain maps from Figure 8 for $\alpha_1$ and $\alpha_4$, and Lemma 2.8 for $\alpha_2$ and $\alpha_3$. We use the usual chain maps when the horizontal strand lies behind the others, and the mirror image chain maps when it is in front. This description comes in three steps; first the underlying surfaces, ignoring disorientation data, then any associated coefficients, and finally the arrangement of disorientation seams. The underlying surface for each map is simply the cylinder over the initial (and final) resolution. Figure 8 shows that $\alpha_1$ carries no coefficient, while $\alpha_4$ carries a coefficient of $\omega^2$.

For the R3 coefficients, notice that the R3 moves occurring in this configuration are one of the following pairs: hlm/↺, mlh/↻, $\overline{R3_{mhl}}^{-1}/\overline{R3_↺}$, or $\overline{R3_{lhm}}^{-1}/\overline{R3_↻}$. Our computation involves an extreme resolution on the $\mathcal{O}$ layer in both R3 moves; let $\sigma_2$ and $\sigma_3$ be the coefficients on the appropriate height (either high or low) components of the $\mathcal{O} \to \mathcal{O}$ part of the corresponding R3 moves. Then, according to Lemma 2.8 and Corollary 2.10, it is always the case that one of the $\sigma_i$'s is $-1$, while the other is $\omega^2$.

In each of the four cases, a (cancelling) pair of reordering signs is needed.

Finally, we add disorientation seams:
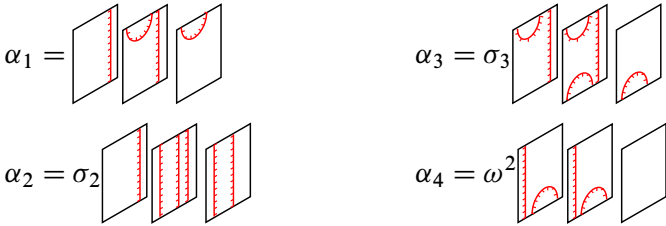


Notice that in $\alpha_2$, the left-most seams on the second two sheets are vertical, as the associated crossing is not involved in the R3 move; the other seams are the unique minimal ones connecting the remaining eight disorientation marks. Similarly, in $\alpha_3$ the right-most seams on the first two sheets are vertical, leaving the others to be determined by minimality.

Thus, our composition $\alpha_4 \circ \alpha_3 \circ \alpha_2 \circ \alpha_1$ is just
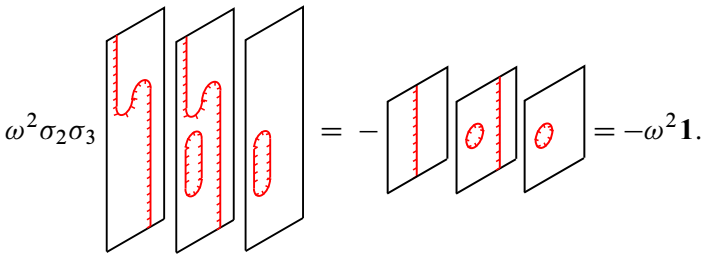


Of course, starting with an extreme object also guarantees this $\alpha$ composition is the only one we need to worry about, as $\beta = 0$ from Lemma 2.9.

The argument for the case in which the left vertical strand is oriented upward, and the right downward, is essentially the same.

Interleaved variations: There are eight variations, and essentially two distinct computations will cover them all. Let's start with hml$^{-1}$/ ↻, ↺ /lmh$^{-1}$, mlh$^{-1}$/mhl, and lhm/hlm$^{-1}$; we'll show the calculation for the first, and explain the necessary

alterations for the other three versions.
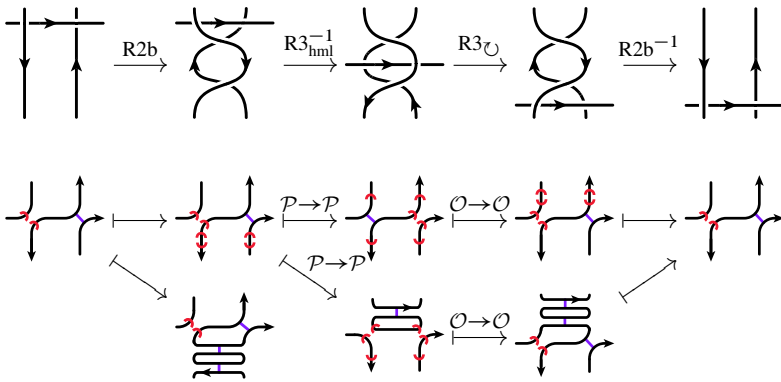


Notice that our first R3 map is ordered $\mathcal{O} \to \mathcal{P}$ and the second $\mathcal{P} \to \mathcal{O}$, each with the high crossing resolved, and that the maps for these moves are labeled by their source and target layers; in particular, the initial $\mathcal{O}$ layer for the second move and the final $\mathcal{P}$ layer for the first move coincide.

Lemma 2.7 tells us there are only three compositions we need to keep track of here. The first map into the second row has an extreme target in the initial $\mathcal{P}$ layer of $R3_{\mathrm{hml}}^{-1}$, which thereafter maps to zero by Lemma 2.9. The composition including the rest of the second row contains a sphere; this is because, disregarding coefficients and disorientation seams, the first and third maps are cylinders from the R2b chain map definitions and Lemma 2.8, the second map contains a cup by Lemma 2.9, and the fourth map, an R2b untuck, contains a cap. Thus we're left with the first row, and a brief look at Lemma 2.8 and Lemma 2.9, and a check that there are no signs from crossing reorderings, confirms that this composition looks like



The calculations for the $\circlearrowright / \mathrm{lmh}^{-1}$, $\mathrm{mlh}^{-1}/\mathrm{mhl}$, and $\mathrm{lhm}/\mathrm{hlm}^{-1}$ variations are very similar. For $\circlearrowright / \mathrm{lmh}^{-1}$, the initial object will have a disoriented left crossing and an oriented right crossing, and we'll resolve each R3 move into layers using the low crossing. Thus we'll need to compute using the mirror image maps, which will introduce an extra factor of $-\omega^2$. The $\mathrm{mlh}^{-1}/\mathrm{mhl}$ and $\mathrm{lhm}/\mathrm{hlm}^{-1}$ variations are even easier:

we start with the doubly oriented object in each case, and resolve into layers using the high crossings or the low crossings, respectively. Crossing reordering maps are trivial in all three of these additional variations, and the overall coefficient for each is just 1.

The computations for $\circlearrowright$ /hml$^{-1}$, lmh$^{-1}$/ $\circlearrowleft$, mhl/mlh$^{-1}$, and hlm$^{-1}$/lhm are somewhat different; again, we'll explicitly show the first.
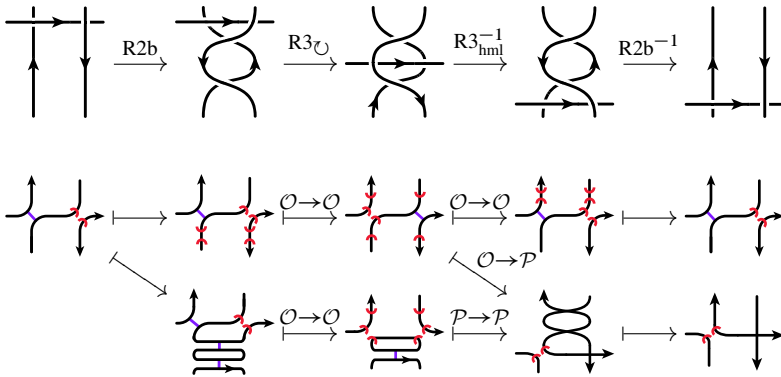


Now our first R3 map is ordered $\mathcal{P} \to \mathcal{O}$ with the high crossing resolved, and the second is ordered $\mathcal{O} \to \mathcal{P}$ with the low crossing resolved. Again, we'll keep track of the layers to which objects belong by referring to the labels on the maps.

By Lemma 2.7, we have three compositions to consider. Two of them factor through the second row, and thus map to a complex with the left crossing disoriented; since our map is a multiple of the identity, these compositions must sum to zero. So we're left with the first row. Using Lemma 2.8 (and its mirror image variant for the second R3), the R2b map definitions, and the fact that crossing reorderings give a minus sign here, it's straightforward to verify this composition is given by



There are a few modifications necessary for lmh$^{-1}$/ $\circlearrowleft$, mhl/mlh$^{-1}$ and for hlm$^{-1}$/lhm. In the lmh$^{-1}$/ $\circlearrowleft$ case, we start with the object with oriented left crossing and disoriented right crossing, and resolve the first R3 on low and the second on high. A crossing reordering sign gives us an overall coefficient of $-\omega^2$. For each of mhl/mlh$^{-1}$ and hlm$^{-1}$/lhm our initial object will be the doubly oriented one, so crossing reordering

maps act trivially. In the calculations, $\text{hlm}^{-1}$ and $\text{mlh}^{-1}$ should be resolved on low, while mhl and lhm should be resolved on high. An overall coefficient of 1 will result in each of these cases.

**MM7**
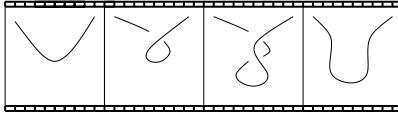


We need to consider four variations of MM7, depending on the orientation of the strand, and whether the "first" crossing is positive or negative. It's easy to check that reversing orientations in the two subsequent calculations doesn't change the result.

First we deal with a positive crossing:
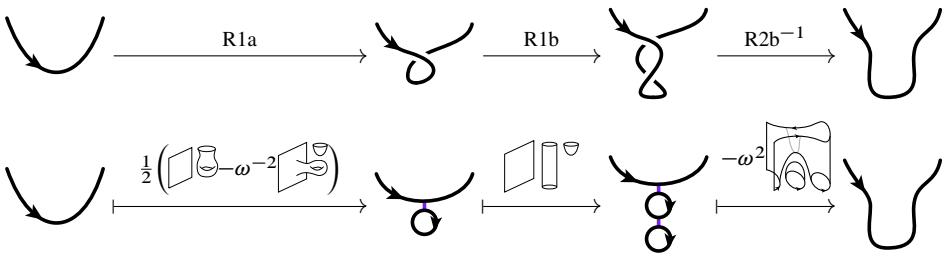


Composing, we see that the second term of the first map gives zero when composed with the later maps. Cancelling the factor of $\frac{1}{2}$ with the torus, we get $-\omega^2$ times the identity.

For the negative crossing, we have
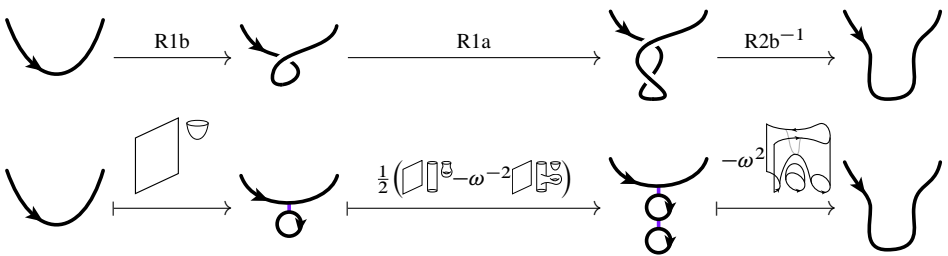


and the composition is just the identity.

**MM8**



This is the only movie move involving all three Reidemeister moves. There are quite a few variations. By a rotation of the whole diagram, we can assume the R1 move happens on the horizontal strand, beginning on the right. Moreover, we can assume that the horizontal strand is oriented right to left (otherwise, we can obtain this condition by a $\pi$ rotation of its time reversal).

There are then sixteen variations, depending on whether the vertical strand lies above or below the horizontal strand, its orientation, the sign of the crossing introduced by the first Reidemeister move in the first frame, and finally whether the first Reidemeister move introduces a twist on the left or right side. The following diagram shows all the maps involved, independent of crossing sign choices and thus without disorientation marks (we will add them later):



(3-2)
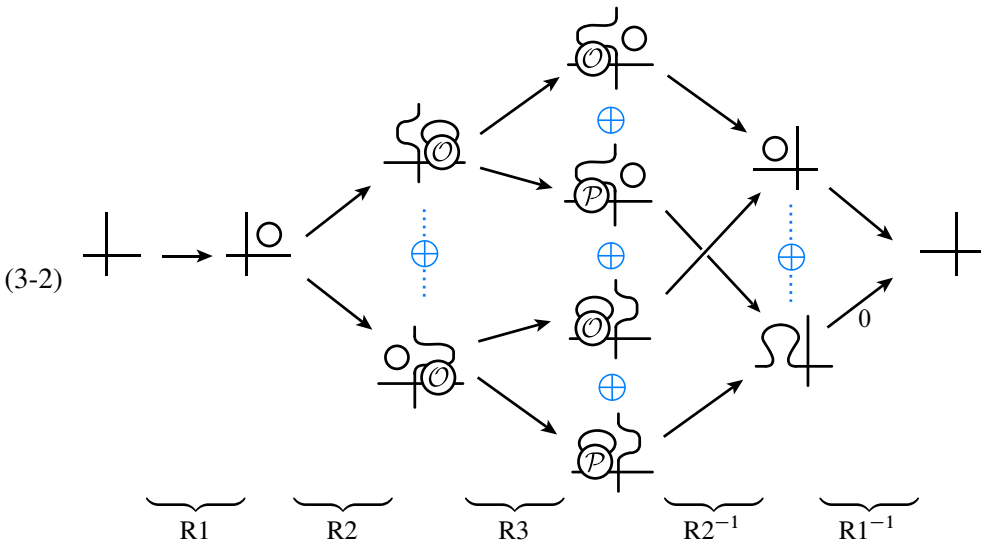
Note that the crossing introduced by the R1 move is always either the low or high crossing in the R3 move, so we will denote its resolution with either $\mathcal{O}$ or $\mathcal{P}$ as we did in the computation for MM6. We can also observe that any map factoring through the resolution
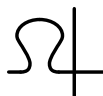
must be zero, since this object maps to zero under R1 (see Section 2.3.2). Thus we need only concern ourselves with the other two compositions in Equation (3-2).

Let's first treat the positive twist. We'll show calculations for the case in which the vertical strand is oriented downward (but ignore whether the twist appears on the left or right side of the horizontal strand; this barely changes any of the calculations). Also, our computation will work regardless of whether the vertical strand is above or below the horizontal strand.

Beginning with a downward-oriented vertical strand, the two relevant compositions are
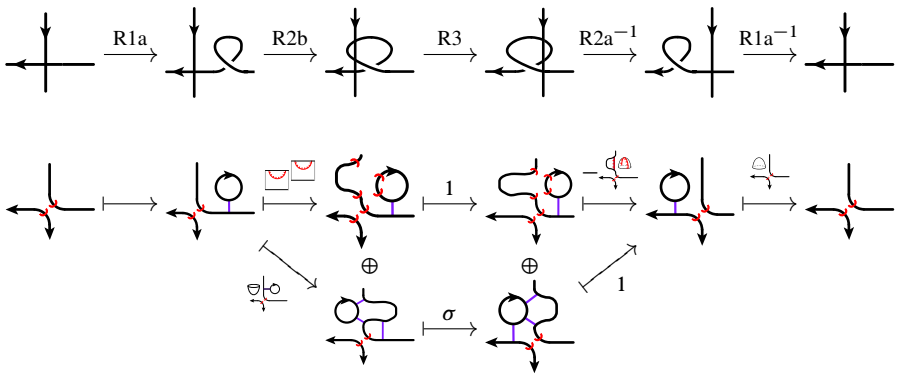


where the initial R1a map is



Here the R3 move is either $\overline{R3_{mlh}}$ or $R3_{hml}^{-1}$, depending on whether the vertical strand is in front or behind.

These chain map components come from Section 2.3.2. In particular, we use Lemma 2.8 (or its "mirror image" analogue, depending on whether the vertical or horizontal strand is on top), to see what the R3 maps do. In the first row our object lies in the $\mathcal{O}$ layer, and thus maps via the identity. (Note that for $\overline{R3_{mlh}}$, when looking for a description of the $\mathcal{O} \to \mathcal{O}$ layer in Lemma 2.8, we actually need to look at the case corresponding to $R3_{mhl}$, since $\overline{R3_{mlh}}$ is defined in terms of $R3_{mhl}$.) All we need to know about the R3 map in the second row, labelled by $\sigma$, is that, ignoring disorientation seams it is the cylinder cobordism. This tells us that the lower row has a spherical

component and can thus be ignored. As such, the composition simplifies to

$$\text{(diagram)} \circ - \text{(diagram)} \circ \mathbf{1} \circ \text{(diagram)} \circ \frac{1}{2} = -\omega^2 \frac{1}{2} \text{(diagram)}$$

$$= -\omega^2 \text{(diagram)}.$$

When the first Reidemeister move introduces a negative crossing, we see instead

$$\text{(diagram)} \xrightarrow{\text{R1b}} \text{(diagram)} \xrightarrow{\text{R2b}} \text{(diagram)} \xrightarrow{\text{R3}} \text{(diagram)} \xrightarrow{\text{R2a}^{-1}} \text{(diagram)} \xrightarrow{\text{R1b}^{-1}} \text{(diagram)}$$

$$\text{(diagram)} \longmapsto \text{(diagram)} \longmapsto \text{(diagram)} \xrightarrow{\gamma} \text{(diagram)} \longmapsto \text{(diagram)} \longmapsto \text{(diagram)}$$
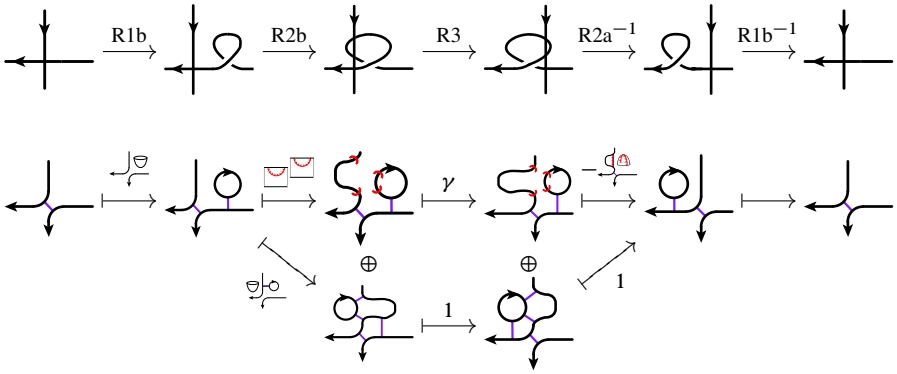
where the final $\text{R1b}^{-1}$ map is

$$\text{(diagram)} \xrightarrow{\frac{1}{2}\left(\text{(diagram)} - \omega^{-2}\text{(diagram)}\right)} \text{(diagram)}.$$

Here the R3 move is either $\overline{\text{R3}_{\text{lmh}}^{-1}}$ or $\text{R3}_{\text{mhl}}^{-1}$, depending on whether the vertical strand is in front or behind.

This time the first row gives zero ($\gamma$ is some disoriented cylinder, so the composition contains a sphere), and we obtain

$$\frac{1}{2}\left(\text{(diagram)} - \omega^{-2}\text{(diagram)}\right) \circ \mathbf{1} \circ \mathbf{1} \circ \text{(diagram)} \circ \text{(diagram)} = \frac{1}{2}\text{(diagram)}$$

$$= \text{(diagram)}.$$

Changing the orientation of the vertical strand (for either a positive or negative twist) alters the computations only slightly (in particular the coefficient appearing on the R3

map is still always $+1$), and we obtain the same result: the coefficient, $1$ or $-\omega^2$, just depends on the sign of the crossing introduced by the first Reidemeister move.

**MM9**
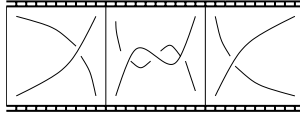


For MM9 we have to be particularly careful; the proof of Proposition 3.1 relied on this movie move, so while checking MM9 we don't have access to any results about the space of chain maps being one dimensional. Thus we'll fully calculate the map, checking it's the identity on every object in the complex associated to the initial tangle.

There are four variations of MM9; we can fix the orientation of one strand, then have to deal with either orientation of the other strand, and either sign for the crossing.

We'll do the calculations for both types of crossings, in a given orientation. It's easy to see that changing an orientation essentially interchanges these cases.

With a positive crossing, we have
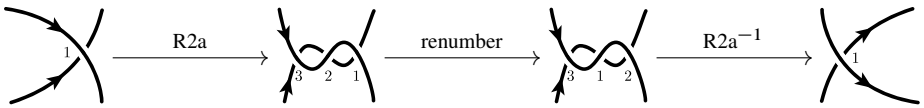


and the components of the chain map are given by:
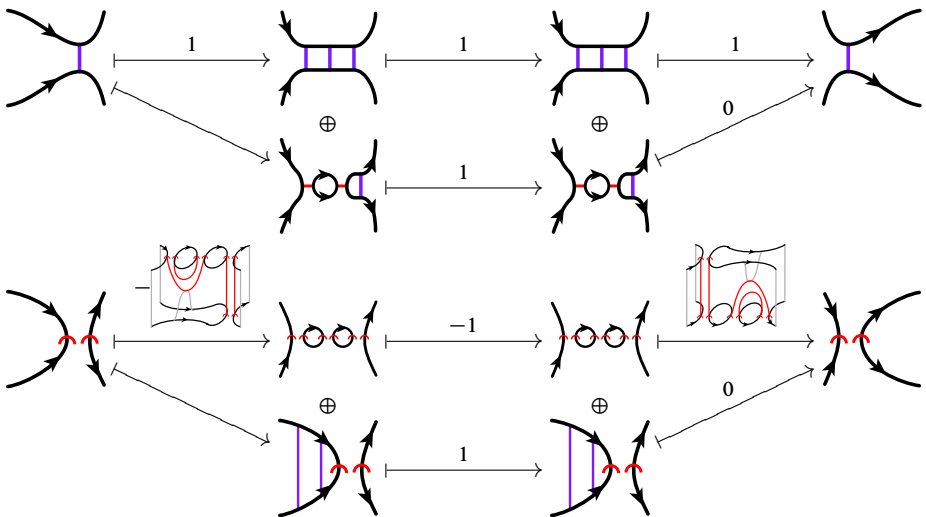
and the composition is just the identity.

With a negative crossing, we have
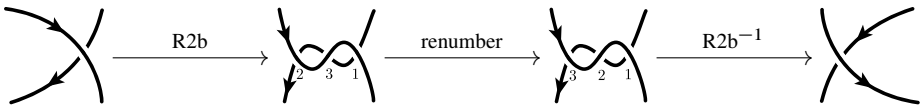


with the components of the chain map being given by
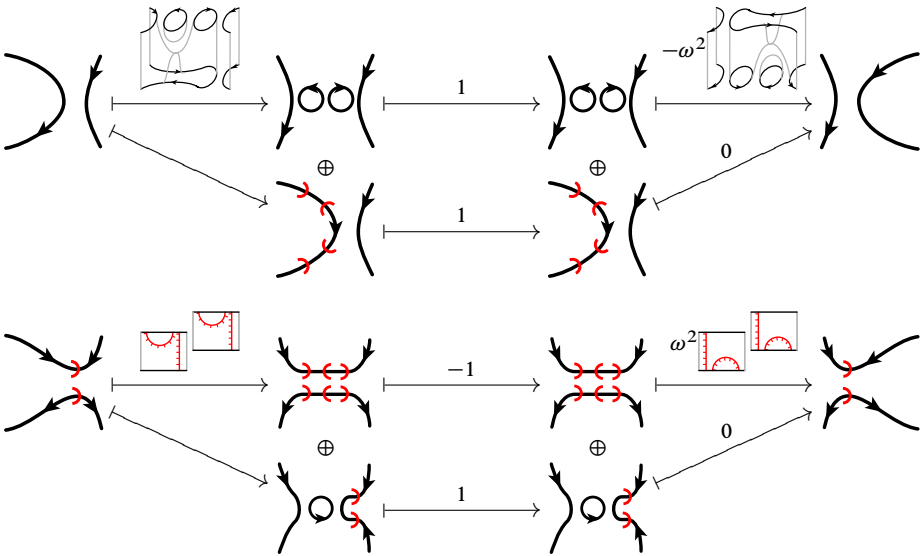


and the composition is $-\omega^2$ times the identity.

## MM10
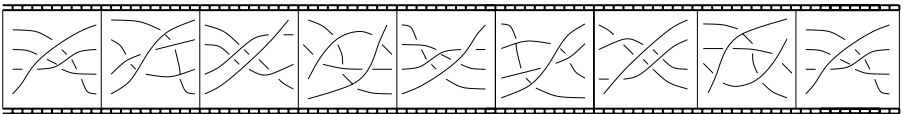


This is the tetrahedron move, and is surprisingly easy. On the other hand, there are a great many variations which we need to treat.

Firstly, let's consider the case in which all strands are oriented to the right. Here, all the crossings are positive, and if we consider the object in the initial complex with homological height zero (ie, we've smoothed every crossing in the oriented way), we see that it is homotopically isolated.

Notice further that each of the eight R3 moves in this movie is of type $R3_{lmh}$, and so has homological ordering $\mathcal{P} \to \mathcal{O}$ when resolving the highest (or lowest) crossing. The oriented smoothing lives in the $\mathcal{O}$ layer, and thus maps via the identity to the oriented smoothing in the next frame by Lemma 2.8. Also, by Lemma 2.7, there is no map to the $\mathcal{P}$ layer. It's easy to see that the same happens at each of the seven other R3 moves, so we're just left with a string of identity maps:
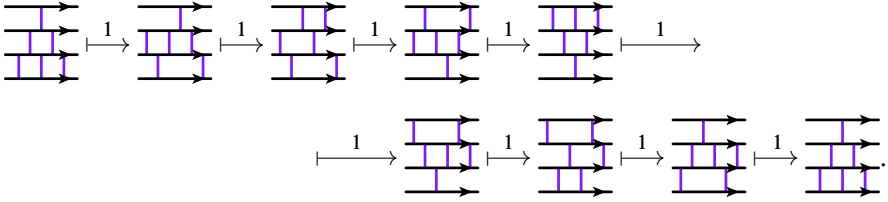


Thus this movie induces the identity chain map.

Beyond this, there are a frightening forty-eight variations. In the space of tangle diagrams, MM10 corresponds to a codimension 2 stratum, appearing as a nongeneric projection in which four strands cross at a point. (See Figure 13). Rotating the projection to put the highest strand in a standard position, there are then 3! height orderings we need to consider for the other strands, and $2^3$ orientations.
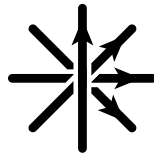


Figure 13: A nongeneric projection corresponding to a MM10 2–cell

It turns out, however, that every variation of MM10 is actually equivalent, modulo MM6.

The idea, essentially, is to add an extra crossing to MM10. We can do this at any adjacent pair of boundary points; for concreteness, let's imagine adding an extra crossing at the top right, with opposite sign to the crossing that already appears in the top right in the first and last frames. There's now a pair of strands carrying two crossings. We can now consider two different variations of MM10, each of which involves only one of those two crossings, and see that these two MM10 moves differ by some MM6 moves (and some "distant Reidemeister moves commute" moves).

More generally, we can stratify the space of smooth tangles so that in the dual cell complex (where a $k$–cell corresponds to a codimension $k$ stratum):

- 0–cells correspond to tangles whose projection to $B^2$ is a generic immersion.
- 1–cells correspond to Reidemeister moves.
- 2–cells correspond to movie moves and pairs of distant Reidemeister moves.
- 3–cells correspond to redundancies amongst movie moves.

If we consider a 3–cell dual to the nongeneric projection shown in Figure 14, we find that the 2–cells on its boundary consist of two MM10 2–cells, four MM6 2–cells, and six distant R-move 2–cells: see Figure 15. Thus invariance for one of the two MM10's, plus invariance for all MM6's and pairs of distant Reidemeister moves, implies invariance for the other MM10.
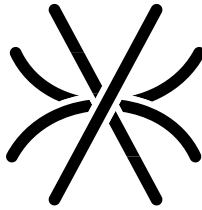


Figure 14: A nongeneric projection corresponding to a 3–cell involving MM10 and MM6

This argument shows that a certain pair of variations of MM10 are equivalent. Thinking about the nongeneric projection corresponding to MM10 in Figure 13, the two variations are related simply by rotating one strand past an adjacent one. It's relatively straightforward to see that these pairs suffice to connect any two variations.

### 3.2.3 MM11–15

**MM11**
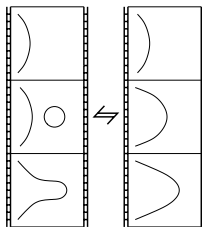


This is trivial in either time direction; the complexes involved only have a single object, and the relevant pairs of cobordisms are isotopic.
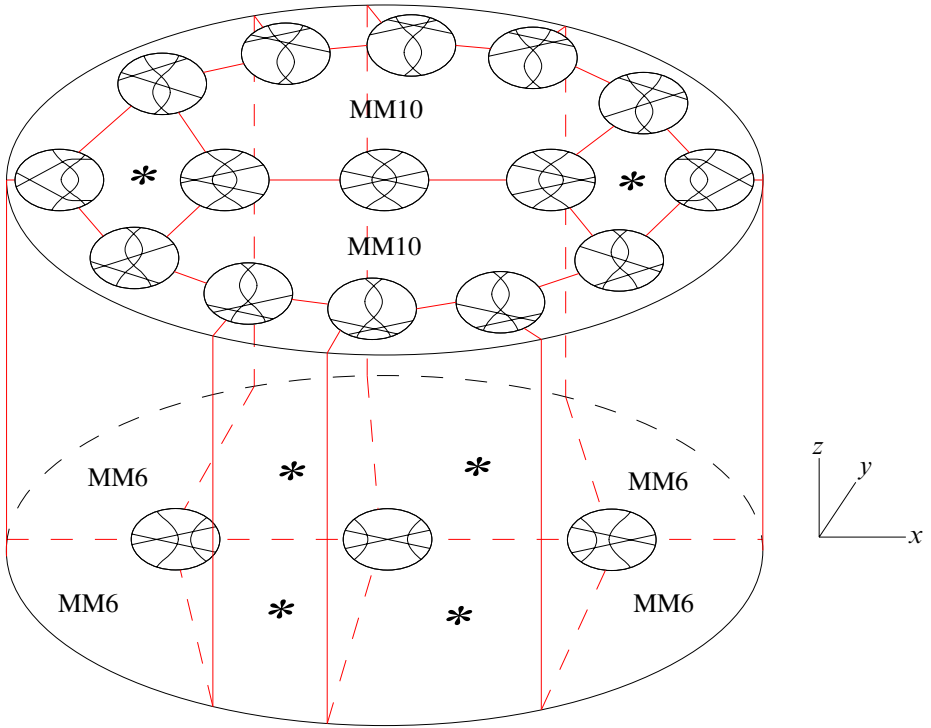
Figure 15: The 3–cell for the singularity in Figure 14, rotated 90 degrees. The 0–cells here are the generic tangle projections neighboring this singularity, achieved by untucking the curved strands ($z$ direction) and translating the crossing ($x$ and $y$ directions). The 2–cells marked with an asterisk correspond to distant Reidemeister moves.
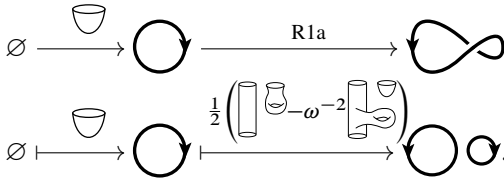
**MM12**
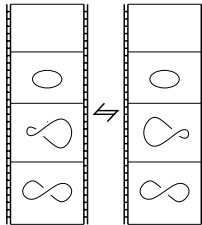


We can't use a homotopy isolation argument for MM12, but it's easy enough to look at all components of the map.

We need to deal with MM12 in two mirror images. In the first mirror image, there is a positive crossing. Reading down, we have on the left

$$\varnothing \longrightarrow \bigcirc \xrightarrow{\text{R1a}} \infty$$

$$\varnothing \longmapsto \bigcirc \xrightarrow{\frac{1}{2}\left(\begin{array}{c}\phantom{x}\end{array} - \omega^{-2}\begin{array}{c}\phantom{x}\end{array}\right)} \bigcirc\bigcirc$$

while on the right we have

$$\varnothing \longrightarrow \bigcirc \xrightarrow{\text{R1a}} \infty$$

$$\varnothing \longmapsto \bigcirc \xrightarrow{\frac{1}{2}\left(\begin{array}{c}\phantom{x}\end{array} - \omega^{-2}\begin{array}{c}\phantom{x}\end{array}\right)} \bigcirc\bigcirc.$$
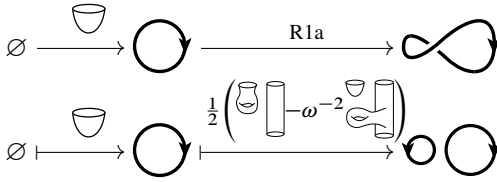
Composing, we see the morphisms agree in the disoriented theory, when $\omega^2 = -1$, but differ by a sign in the unoriented theory.

Reading up, we have on the left

$$\varnothing \longleftarrow \bigcirc \xleftarrow{\text{R1a}^{-1}} \infty$$

$$\varnothing \longleftarrow \bigcirc \longleftarrow \bigcirc\bigcirc$$

$$\oplus$$

$$0 \xleftarrow{0} \infty$$

while on the right we have

$$\varnothing \longleftarrow \bigcirc \xleftarrow{\text{R1a}^{-1}} \infty$$

$$\varnothing \longleftarrow \bigcirc \longleftarrow \bigcirc\bigcirc$$

$$\oplus$$

$$0 \xleftarrow{0} \infty.$$
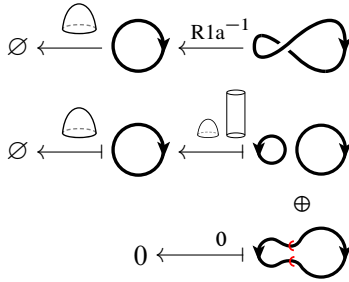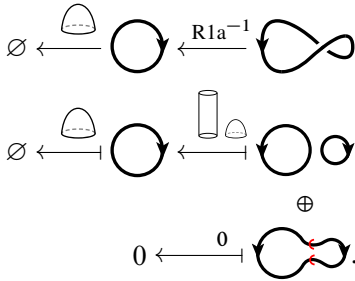
These chain maps agree exactly, in both the unoriented and the disoriented theory.

The mirror image is much the same, although it's the forward in time maps that agree exactly, and the backwards in time maps that agree up to a factor of $-\omega^2$.

**MM13**
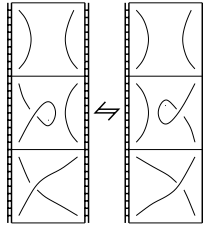


This time there are no orientation variations; we can take both strands in the initial frame to be oriented upwards. We need to compare the two clips read both up and down, and also consider the mirror image.

Reading down we have on the left
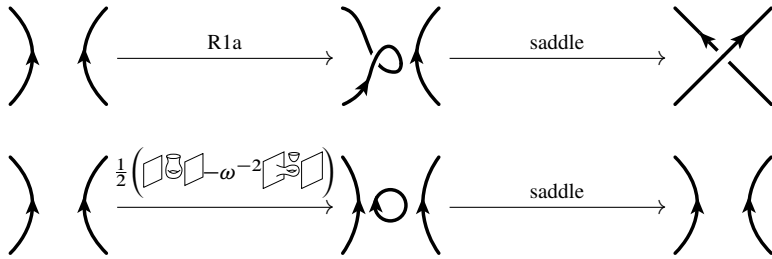


and on the right
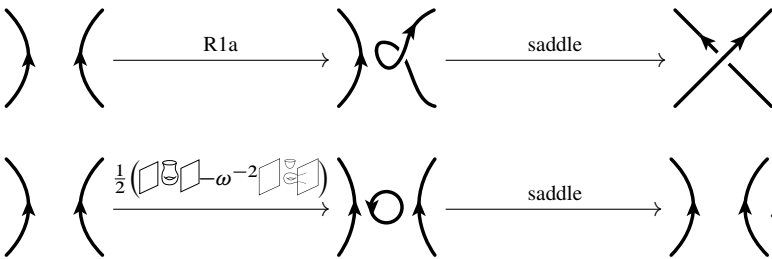


These maps differ by a sign of $-\omega^2$. Reading up, both maps are the identity on the oriented smoothing, and zero on the disoriented smoothing, and hence agree on the nose.

In the mirror image, we see the opposite pattern (since the "interesting" morphism in the R1a and R1b maps appears in opposite directions). There are no other orientations to deal with.

**MM14**
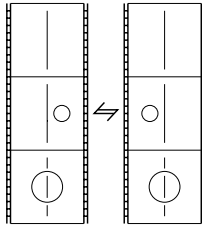


Fixing the orientation of the strand to be from bottom to top, the loop can either be clockwise or counterclockwise, and lie either below or above the strand. We'll first deal with the case in which is loop is oriented counterclockwise, and lies below the strand.

We can't use a homotopy isolation argument for MM12, but it's easy enough to look at all components of the map. Because we're looking at all components, we actually need to pay attention to the ordering of crossings; for compatibility with the Reidemeister maps described in Section 2.3.3, we'll number the crossings from the bottom up, so that the negative crossing comes first.

On either the left or right sides of MM14, we have an R2 map. Looking at Figure 7 and Figure 8, we see that on both sides we obtain the map:
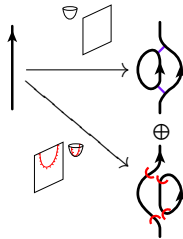


Backwards in time, we obtain different maps. On the left we see
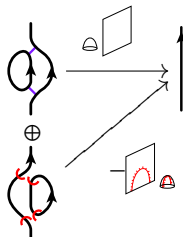
and on the right:
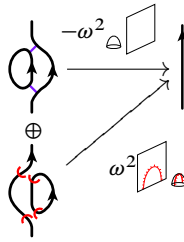


Thus we see that forwards in time the maps agree, but backwards in time they only agree in the disoriented theory.

Reversing the relative heights of the loop and the strand doesn't change the calculation; similarly reversing the orientation of one strand has no effect.

**MM15**   We now consider both time directions in MM15.
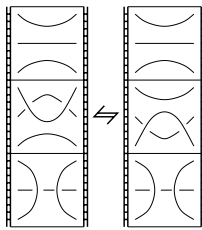


We need to deal with 4 variations; assuming the middle strand is oriented left to right, we can orient the highest strand either to the left or to the right (forcing the lowest strand to be oriented oppositely), and we can tuck the middle strand either under or over the other strands.

We'll start by choosing orientations so the upper two strands are oriented to the right, and the lowest strand is oriented to the left, and tuck the middle strand under the others.
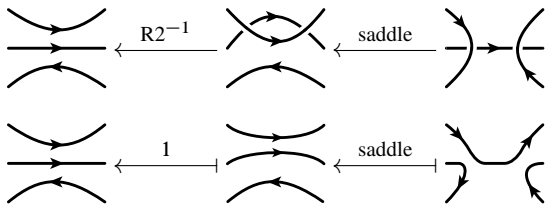
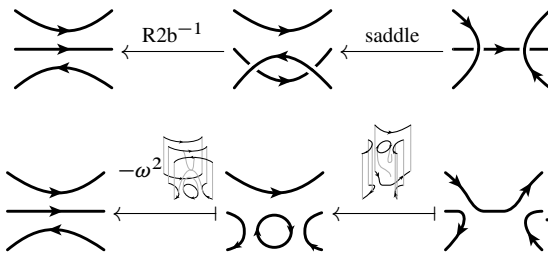Reading down, we have on the left
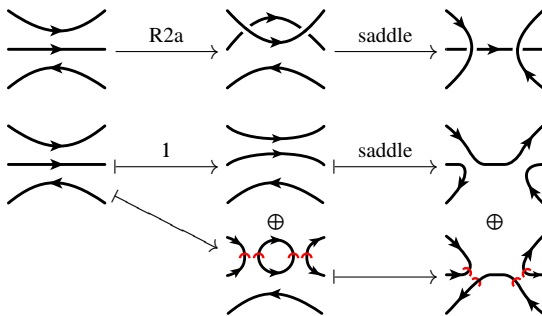
and on the right
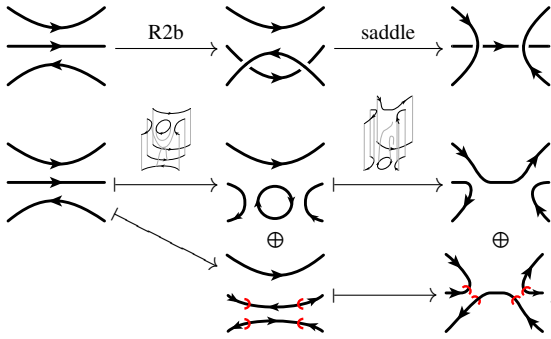


We've left some maps we don't need to know about unlabeled.

Looking only at the component of the maps going to $\overset{\frown}{\supset}\underset{\smile}{\varsigma}$, we see each side of the movie move agrees on the nose; both maps are a saddle involving the lower two strands.

Reading up, we have on the left



and on the right



We see that the two movies differ by a sign of $-\omega^2$.

The other variations turn out exactly the same way. Changing the orientations of the highest and lowest strand has no effect; we simply interchange R2a and R2b maps throughout. Switching the height ordering interchanges R2al with R2ar, and R2b+ with R2b−, with no net effect.

This concludes the proofs of Theorem 1.1 and Theorem 1.2.

# 4 Odds and ends

## 4.1 Recovering Jacobsson's signs

Summarizing the results of the above calculations at $\omega = 1$ (ie in the original unoriented theory), in Table 2, we see that in most cases we agree with the signs Jacobsson observed [10]. There are exceptions, however (shown highlighted in the tables).

In particular, MM6 (Jacobsson's number 15) does not appear to exhibit a sign problem in the unoriented theory (whereas we see both $-\omega^2$ and 1 in different variations), and the two mirror images of MM12 (Jacobsson's number 12) both exhibit a sign problem, one forwards in time, one backwards. These disagreements coincide with calculations performed by the first author using Lee's [19] variant of Khovanov homology. Further, we can't easily extract the sign in the unoriented theory for all the variations of MM10 (without analysing which variants of MM6 are used in each of the 3–cells shown in Figure 15), so we don't know whether these agree with Jacobsson's values.

| MM | J♯ | ± |
|---|---|---|
| 6 | 15 | + |
| 7 | 13 | - |
| 7 (mirror) | 13 | + |
| 8 | 6 | - |
| 8 (mirror) | 6 | + |
| 9 | 14 | + |
| 9 (mirror) | 14 | - |
| 10 | 7 | + |

| MM | J♯ | ↓ | ↑ |
|---|---|---|---|
| 11 | 9 | + | + |
| 12 | 11 | - | + |
| 12 (mirror) | 11 | + | − |
| 13 | 12 | - | + |
| 13 (mirror) | 12 | + | - |
| 14 | 8 | + | - |
| 15 | 10 | + | - |

Table 2: The signs observed in the unoriented theory

## 4.2 Relationship with the unoriented invariant

In this section we'll prove that for knots and links (that is, ignoring tangles and cobordisms), the disoriented and unoriented invariants are equivalent. We'll write $[\![L]\!]_D$ and $[\![L]\!]_U$ for the disoriented and unoriented invariants of $L$ respectively.

**Theorem 4.1** *There is a fully faithful functor* Alt: $\mathbf{UnAb}_0 \otimes \mathcal{R}[\omega] \hookrightarrow \mathbf{DisAb}_0$ *(the subscript 0 denotes the part of the canopolis with no boundary points), which "alternately orients" each unoriented diagram. This induces a functor* Alt: $\mathrm{Kom}(\mathbf{UnAb})_0 \otimes \mathcal{R}[\omega] \hookrightarrow$ $\mathrm{Kom}(\mathbf{DisAb})_0$ *such that*

$$\mathrm{Alt}\,([\![L]\!]_U) \cong [\![L]\!]_D,$$

*although this isomorphism isn't canonical.*

**Proof** We've already seen the forgetful map **DisAb** → **UnAb**, setting $\omega = 1$ and forgetting orientation data. It's relatively easy to see that this guarantees that we can reconstruct the unoriented invariant from the disoriented one (for tangles too!). To see the two invariants are actually equivalent, we'll introduce a new canopolis of "alternately oriented cobordisms", **AltAb**, a subcanopolis of **DisAb**. We'll construct an isomorphism $\mathbf{UnAb}_0 \otimes \mathcal{R}[\omega] \cong \mathbf{AltAb}_0$, and additionally show that the invariant of a knot or link (but not a tangle!), which is an up-to-homotopy complex in $\mathbf{DisAb}_0$, always has a representative in the subcategory $\mathbf{AltAb}_0$, which coincides with the image of the unoriented invariant in $\mathbf{AltAb}_0$.

The category $\mathbf{AltAb}_0$ consists of diagrams in the disc comprised of oriented loops, such that all "outermost" loops are oriented counterclockwise, and at each successive depth of nesting, the orientations reverse. This is a subset of the objects of $\mathbf{DisAb}_0$. The morphisms of $\mathbf{AltAb}_0$ are simply all the morphisms of $\mathbf{DisAb}_0$ between these objects. In fact, $\mathbf{AltAb}_0$ is the "boundaryless" part of a full canopolis **AltAb** defined in much the same way.

The isomorphism $\mathbf{UnAb}_0 \otimes \mathcal{R}[\omega] \cong \mathbf{AltAb}_0$ is easy; simply orient the circles in an object of $\mathbf{UnAb}_0$ in the prescribed manner, and note that for any cobordism, these orientations always extend to an honest orientation of the cobordism. It's injective by Lemma 2.1. It's surjective because every cobordism in $\mathbf{AltAb}_0$, which *a priori* might have disorientation seams, is actually a power of $\omega$ multiple of a properly oriented cobordism, by the following two Lemmas.

**Lemma 4.2** *Reversing the fringe of a closed disorientation seam gives a sign of $-1$.*

**Proof** Use the neck cutting relation parallel to the seam. $\square$

**Lemma 4.3** *If $Y$ is a disoriented surface with all disorientation seams closed, and with alternately oriented boundary components, then $Y$ is equal to a power of $\omega$ multiple of the homeomorphic oriented surface.*

**Proof** By applying fringe moves, we can assume that the disorientation seam is connected on each connected component of $Y$. (If necessary, reverse fringe directions using the previous lemma.) The assumption about boundary orientations now implies that the seam is null-homologous, and so can be removed via further fringe moves. $\square$

We next discover how to push a link complex $[[L]]$ in $\mathbf{DisAb}_0$ down into the subcategory $\mathbf{AltAb}_0$.

We begin with a quick statement about the disorientations that can appear on a circle.

**Lemma 4.4** *Define the "disorientation number" of a disoriented circle to be the number of counterclockwise facing disorientation marks minus the number of clockwise facing disorientation marks. (See Figure 16.) Then two disoriented circles $C_1$ and $C_2$ are isomorphic in* **DisAb** *exactly if their disorientation numbers agree.*
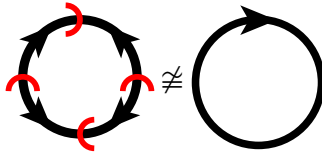


Figure 16: A disoriented circle with disorientation number $+2$ is not isomorphic in **DisAb** to an oriented circle.

We want to show that every circle appearing in an object of $[\![L]\!]$ has disorientation number $0$. This is a conservation argument; near each disorientation mark on the circle, there used to be a crossing in $L$, either just inside or outside the circle. Whether the disorientation mark faces counterclockwise or clockwise records whether the two strands in the crossing were oriented "inwards" or "outwards" across the circle. Since the original link must cross any given circle a total of $0$ times, the signed count of disorientation marks is $0$ as well. This shows that every object appearing in $[\![L]\!]$ is actually isomorphic to the corresponding object appearing in $[\![L]\!]_U$ (this is Lemma 4.4). In fact, the only choice in this isomorphism is a multiple of $\pm 1$ or $\pm \omega$.

Thus we take the link complex $[\![L]\!]$, and replace every disoriented circle with the appropriately oriented circle. The complex now lies entirely within the subcategory **AltAb$_0$**. This complex agrees with the unoriented link complex, thought of as living in **AltAb$_0$**, except for the fact that each morphism may be off by a unit, simply because the underlying surfaces for each morphism are the same, and by Lemma 4.3 above the morphisms are unit multiples of each other.

A little combinatorial lemma about sprinkling units in a complex gets us to the desired result.

**Lemma 4.5** (Sprinkling units) *Suppose we have two anticommutative cubes, with identical objects, such that corresponding morphisms only ever differ by a unit. Further suppose that the composition of any two "edges" of the cube is nonzero. Then the two cubes are isomorphic, via a map which just multiplies each object in the cube by some unit.*

**Remark** The hypothesis that the composition of any two composable maps in the cube is nonzero certainly holds in the case we're interested in. The complex associated to a knot or link has as morphisms pairs of pants (or their time reverse) and cylinders, and it's easy to see that any way of composing two pairs of pants is nonzero.

**Remark** Something like this lemma is used in [15] in describing a categorification of the colored Jones polynomial, without the need for functoriality. Note also that our construction of a properly functorial version of Khovanov homology should make a more direct construction of a categorification of the colored Jones polynomial possible, and allow the possibility of this categorification itself being functorial. See page 20 of [15].

**Proof** An easy induction on the dimension of the cube. For one dimensional cubes, the result is trivial. For any cube, by induction we can choose an isomorphism $\phi_t$ between the top layers of the cubes, and another $\phi_b$ between the bottom layers of the cubes. Now we need to tweak the top layer isomorphism, so together the isomorphisms give an isomorphism on the entire cube. Consider the "highest" vertical differential $d_v$, between the initial objects in the top and bottom layers, and define a unit $\epsilon$ by $d_v\phi_t = \epsilon\phi_b d_v$. Now replace the isomorphism $\phi_t$ with $\epsilon\phi_t$. We now just need to check that our isomorphism $\phi$ commutes with every vertical differential. Thus consider a square of differentials in one cube

$$
\begin{array}{ccc}
\bullet & \xrightarrow{\ d_t^1\ } & \bullet \\
\Big\downarrow{\scriptstyle d_l^1} & \scriptstyle d_b^1 & \Big\downarrow{\scriptstyle d_r^1} \\
\bullet & \xrightarrow{\ \ \ } & \bullet
\end{array}
$$

with $d_t^1$ a differential in the top layer, and $d_b^1$ a differential in the bottom layer. There's a corresponding square of differentials in the other cube, with differentials $d_t^2, d_b^2, d_r^2$ and $d_l^2$. By our construction $\phi d_t^1 = d_t^2\phi$, and $\phi d_b^1 = d_b^2\phi$, and we'll assume further $\phi d_l^1 = d_l^2\phi$ (we're going to apply this piece of the argument to every such square, starting with $d_l = d_v$, the "highest" vertical differential described above). Now we know $\phi d_r^1 = \zeta d_r^2\phi$ for some unit $\zeta$; we just need to show $\zeta = 1$. We then deduce the following equations

$$
\begin{aligned}
\phi d_r^1 d_t^1 &= \zeta d_r^2 \phi d_t^1 & \phi d_b^1 d_l^1 &= \zeta d_b^2 d_l^2 \phi \\
&= \zeta d_r^2 d_t^2 \phi, & &= \zeta \phi d_b^1 d_l^1,
\end{aligned}
$$

and, making use of the hypothesis that the composition $d_b^1 d_l^1$ is nonzero, conclude that $\zeta$ is indeed 1. $\qquad\square$

That concludes the proof of Theorem 4.1. $\qquad\square$

## 4.3   Sliding a handle past a crossing

In this section we give an example calculation in the new setup, illustrating an interesting difference with the unoriented construction of Khovanov homology.

Consider the following two cobordisms from ╳ to itself: the first is the identity except for a handle attached to the over-sheet to the right of the crossing, and the second is the same except that the handle is attached to the over-sheet to the left of the crossing. We'll denote these schematically by

$$F = \text{⧖}\quad \text{and}\quad G = \text{⧗}.$$

**Proposition 4.6**   *F and G are homotopic, $F \simeq G$, as maps in* Kom(**DisAb**), *whereas in* Kom(**UnAb**), *we have $F \simeq -G$ instead.*

**Proof**   Note that these cobordisms are clearly isotopic, and so the functoriality result above gives us an automatic proof. (Exercise: figure out the sequence of movie moves relating them!) However, we will construct an explicit homotopy: the arrow marked $h$ in Figure 17.
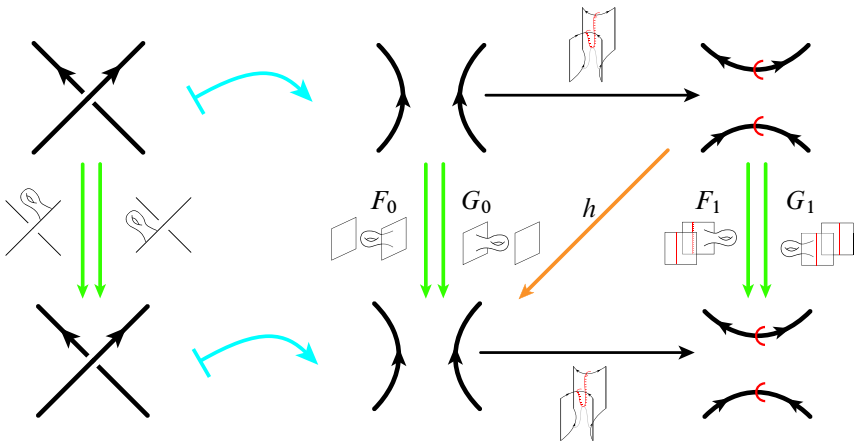


Figure 17: Chain maps and homotopy for the Proposition 4.6

We need a homotopy $h$ such that $hd + dh = F_i - G_i$, and propose

$$h = 2\omega^{-1}\text{�}.$$

At height zero we then have

$$hd + dh = 2\omega^{-1}$$ 

$$= \omega^{-1} \left( \quad + \quad \right)$$ 

$$= \omega^{-1} \left( \omega \quad + \omega^{-1} \quad \right)$$ 

$$= \quad + \omega^{-2} \quad ,$$ 

which is just $F_0 - G_0$ when we set $\omega^2 = -1$. There's a similar computation at height 1:

$$hd + dh = 2\omega^{-1}$$ 

$$= 2$$ 

$$= \quad + \quad$$ 

$$= \quad + \omega^{-2} \quad .$$ 

Again, setting $\omega^2 = -1$ makes the last line $F_1 - G_1$, which gives the result. $\square$

## 4.4 Confusions

In this final section, we'll describe a defect in the discussion so far, and say a little about a proposal to fix it.

The construction we've proposed so far is a functor from the category of oriented tangles, **OrTang**, into the category of complexes of disoriented flat tangles Kom(**DisAb**). In particular, it only gives maps for oriented cobordisms between oriented links. This isn't really ideal; the old unoriented theory gave maps for nonorientable cobordisms. For example, while a Möbius band with positive $\frac{3}{2}$ twists provides a generator of the Khovanov invariant of the trefoil in the old theory, our construction doesn't know what to do with nonorientable surfaces.

Thus we'd like to extend the theory to a functor from **DisTang**, the category of disoriented tangles. On the level of objects, this is no problem; simply map disorientations to disorientations. Unfortunately, there is now an additional Reidemeister move, namely "sliding a disorientation through a crossing", which we'll name a "vertigo", for which we need to provide an isomorphism between the corresponding complexes. Further, we'd need to check additional movie moves, relating this new Reidemeister move to the original three.[3]

However, it's easy to see that it just isn't possible to produce a homotopy equivalence between the corresponding complexes in Kom(**DisAb**). To begin, such a homotopy equivalence would have to be an isomorphism; using Lemma 3.5, we see no homotopies are possible in the complex for a single crossing, regardless of any additional disorientations. Such an isomorphism would presumably be of the form in Figure 18.

In particular, we'd need an isomorphism in **DisAb** reversing the direction of a disorientation mark on a disoriented strand.

Such an isomorphism, which we'll dub a "confusion", would necessarily be a troublesome thing; if the confusion were simply to be some "local" structure on a surface, which I'll draw here as a box labeled by $c$ (or a box labeled by $c^{-1}$ for its inverse), we could perform the calculation



$$(4\text{-}1) \qquad \omega = \quad = \quad = \quad = \omega^{-1}$$

---

[3]There's actually a big incentive for this extension; it turns out that all the different oriented versions of the usual 15 movie moves become equivalent modulo these extra movie moves involving disorientations. This was actually our original motivation for introducing confusions.
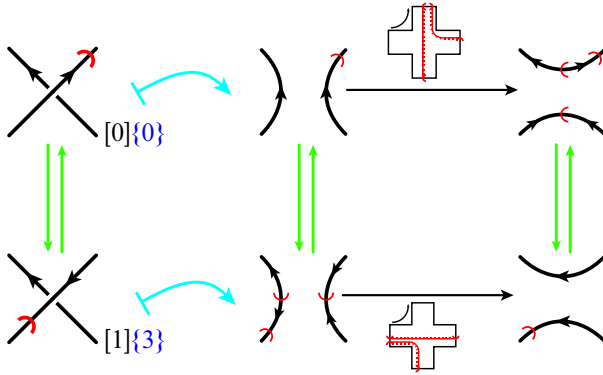
Figure 18: A hypothetical isomorphism of complexes implementing a particular case of the "vertigo" move

producing a contradiction with the requirement that $\omega^2 = -1$.

The way out of this seems to be to make the confusion a spinorial object, so an extra sign gets introduced as we drag the confusion around the circle, in the third equality in Equation (4-1).

At this point it seems appropriate to apologise for having talked about a particular diagrammatic model for such "spinorial confusions" at various conferences, but to be omitting the details in this paper. We still intend to write these details down!

We'll briefly list the improvements to the theory we anticipate being able to make, after the introduction of confusions.

- Connecting the category **DisAb**; in particular, all disoriented circles would be isomorphic.

- Extending the invariant to disoriented tangles, and disoriented cobordisms between them.

- Using the categorified Kauffman trick, to more easily describe the Reidemeister 3 chain map.

- After checking additional movie moves involving vertigos, being able to reduce the computations required in Section 3.2, by taking advantage of the fact that all oriented versions of each oriented movie move become equivalent module disoriented movie moves.

# Appendix

Throughout this appendix, we'll at times just write a "bullet", $\bullet$, for a matrix entry which we don't need to care about.

## A.1 Gaussian elimination

**Lemma A.1** (Gaussian elimination for complexes) *Consider the complex*

$$(A.1) \qquad A \xrightarrow{\left(\begin{smallmatrix} \bullet \\ \alpha \end{smallmatrix}\right)} \begin{matrix} B \\ \oplus \\ C \end{matrix} \xrightarrow{\left(\begin{smallmatrix} \varphi & \lambda \\ \mu & \nu \end{smallmatrix}\right)} \begin{matrix} D \\ \oplus \\ E \end{matrix} \xrightarrow{(\bullet\ \epsilon)} F$$

*in any additive category, where $\varphi\colon B \xrightarrow{\cong} D$ is an isomorphism, and all other morphisms are arbitrary (subject to $d^2 = 0$, of course). Then there is a homotopy equivalence with a much simpler complex, "stripping off" $\varphi$.*



**Remark** Gaussian elimination is a strong deformation retract. In fact, it preserves the simple homotopy type of the complex.

**Proof** This is simply Lemma 4.2 in Bar-Natan [3] (see also Figure 2 there), this time explicitly keeping track of the chain maps. □

We'll also state here the result of applying Gaussian elimination twice, on two adjacent (but noncomposable) isomorphisms. Having these chain homotopy equivalences handy will tidy up the calculations for the Reidemeister 2 and 3 chain maps.

**Lemma A.2** (Double Gaussian elimination) *When $\psi$ and $\varphi$ are isomorphisms, there's a homotopy equivalence of complexes:*



**Proof** Apply Lemma A.1, killing off the isomorphism $\psi$. Notice that the isomorphism $\varphi$ survives unchanged in the resulting complex, and apply the lemma again. □

**Remark** Convince yourself that it doesn't matter in which order we cancel the isomorphisms!

## A.2 Calculations of Reidemeister chain maps

We can now go through the constructions of the Reidemeister chain maps.

**Lemma A.3** *The chain maps displayed in Figures 5 and 6 are homotopy equivalences.*

**Proof** We'll just do the R1a move; the R1b is much the same.

The complex associated to  is



with $d$ simply the disoriented saddle. Delooping at homological height 1, and cancelling the disorientations at height 2, using the isomorphisms

with inverses

$$\zeta_1^{-1} = \left( \begin{array}{cc} \square^{\includegraphics{x}} & \tfrac{1}{2}\includegraphics{x} \end{array} \right), \qquad\qquad \zeta_2^{-1} = \includegraphics{x},$$

we obtain the complex



The differential here is the composition $\zeta_2 d \zeta_1^{-1}$. Stripping off the isomorphism $\varphi$, according to Lemma A.1, we see that the complex is homotopy equivalent to the desired complex: a single strand. The "simplifying" homotopy equivalence is

$$s_1 = \begin{pmatrix} 0 & 1 \end{pmatrix} \circ \zeta_1 = \square_{\includegraphics{x}}, \qquad\qquad\qquad s_2 = 0$$

with inverses

$$s_1^{-1} = \zeta_1^{-1} \circ \begin{pmatrix} -\varphi^{-1}\lambda \\ 1 \end{pmatrix} = \frac{1}{2}\left( \square\includegraphics{x} - \omega^{-2}\includegraphics{x} \right), \qquad s_2^{-1} = 0$$

as claimed. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

**Lemma A.4** *The chain maps displayed in Figures 7 and 8 are homotopy equivalences.*

**Proof** We'll deal with the R2a move first.

The complex associated to $\includegraphics{x}$ is



with differentials

$$d_{-1} = \begin{pmatrix} \includegraphics{x} \\ \includegraphics{x} \end{pmatrix} \qquad\qquad d_0 = \begin{pmatrix} \includegraphics{x} & -\includegraphics{x} \end{pmatrix}.$$

(In these matrix entries, and below, an interval connecting two points on a smoothing indicates a saddle cobordism along that interval.)

Applying the delooping isomorphism

$$\begin{pmatrix} \frac{1}{2\omega} \,\diagup\!\ominus\!\diagdown \\ \diagup\!\ominus\!\diagdown \end{pmatrix},$$

which has inverse

$$\begin{pmatrix} \diagup\!\ominus\!\diagdown & \frac{1}{2\omega} \,\diagup\!\ominus\!\diagdown \end{pmatrix},$$

to the direct summand with a loop, we obtain the complex

$$q^{-1} \,\diagup\!\diagdown\!\diagup\!\diagdown \quad \xrightarrow{d_{-1}} \quad \begin{matrix} \overparen{\phantom{xx}} \\ \oplus \\ q \,\diagup\!\diagdown \\ \oplus \\ q^{-1}\,\diagup\!\diagdown \end{matrix} \quad \xrightarrow{d_0} \quad q \,\diagup\!\diagdown\,\diagdown$$

$$d_{-1} = \begin{pmatrix} \gamma = \,\diagup\!\diagdown \\ \bullet \\ \psi = \,\diagup\!\diagdown \end{pmatrix} \qquad d_0 = \begin{pmatrix} \lambda = \,\bowtie & \varphi = \,\diagdown\!\diagup & \bullet \end{pmatrix}.$$

Here we've named the entries of the differentials in the manner indicated in [Lemma A.2](). Applying that lemma gives us chain equivalences with the desired one object complex. The chain equivalences we're after are compositions of the chain equivalences from [Lemma A.2]() with the delooping isomorphism or its inverse.

Thus the R2a "untuck" chain map is

$$\begin{pmatrix} 1 & 0 & -\gamma\psi^{-1} \end{pmatrix} \circ \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2\omega}\,\diagup\!\ominus\!\diagdown \\ 0 & \diagup\!\ominus\!\diagdown \end{pmatrix} = \begin{pmatrix} 1 & -\,\diagup\!\diagdown\!\circ\,\diagup\!\ominus\!\diagdown \end{pmatrix}$$

as claimed, and the "tuck" map is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \diagup\!\ominus\!\diagdown & \frac{1}{2\omega}\,\diagup\!\ominus\!\diagdown \end{pmatrix} \circ \begin{pmatrix} 1 \\ -\varphi^{-1}\lambda \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ \diagup\!\ominus\!\diagdown\,\circ\,\bowtie \end{pmatrix},$$

Now the R2b move, in much the same way. The complex associated to ⟨diagram⟩ is

$$
q^{-1}\,\text{⟨diagram⟩}\ \Big(\ \xrightarrow{\ d_{-1}\ }\ \text{⟨diagram⟩}\oplus\text{⟨diagram⟩}\ \xrightarrow{\ d_0\ }\ q\ \Big)\,\text{⟨diagram⟩}
$$

with differentials

$$
d_{-1}=\begin{pmatrix}\text{⟨diagram⟩}\\[2pt]\text{⟨diagram⟩}\end{pmatrix}
\qquad\qquad
d_0=\big(-\text{⟨diagram⟩}\ \ \text{⟨diagram⟩}\big)\,.
$$

This time instead of just delooping, we'll also cancel the obvious pairs of disorientation marks. The isomorphism we'll use is

$$
\zeta^{\bullet}=\Big(\omega^{-1}\text{⟨diagram⟩}\Big),\ 
\begin{pmatrix}\omega^{-2}\text{⟨diagram⟩} & 0\\[2pt] 0 & \tfrac12\,\text{⟨diagram⟩}\\[2pt] 0 & \text{⟨diagram⟩}\end{pmatrix},\ 
\Big(\omega^{-1}\text{⟨diagram⟩}\Big),
$$

with inverses

$$
(\zeta^{-1})^{\bullet}=\Big(\text{⟨diagram⟩}\Big),\ 
\begin{pmatrix}\text{⟨diagram⟩} & 0 & 0\\[2pt] 0 & \text{⟨diagram⟩} & \tfrac12\,\text{⟨diagram⟩}\end{pmatrix},\ 
\Big(\text{⟨diagram⟩}\Big)\,.
$$

We obtain the complex

$$
q^{-1}\,\text{⟨diagram⟩}\ \Big(\ \xrightarrow{\ d_{-1}\ }\ q\ \Big)
\ \begin{array}{c}\text{⟨diagram⟩}\\[4pt]\oplus\\[4pt]q^{-1}\text{⟨diagram⟩}\\[4pt]\oplus\end{array}
\ \Big(\ \xrightarrow{\ d_0\ }\ q\ \Big)\,\text{⟨diagram⟩}
$$

where

$$
d_{-1}=\begin{pmatrix}\gamma=\omega^{-1}\text{⟨diagram⟩}\\[2pt]\bullet\\[2pt]\psi=\omega\,\text{⟨diagram⟩}\end{pmatrix}
\qquad
d_0=\big(\lambda=-\text{⟨diagram⟩}\ \ \varphi=\text{⟨diagram⟩}\ \ \bullet\big)\,.
$$

Thus the R2b "untuck" chain map is

$$
\big(1\ \ 0\ \ -\gamma\psi^{-1}\big)\circ
\begin{pmatrix}\omega^{-2}\text{⟨diagram⟩} & 0\\[2pt] 0 & \tfrac12\,\text{⟨diagram⟩}\\[2pt] 0 & \text{⟨diagram⟩}\end{pmatrix}
=\big(\omega^{-2}\text{⟨diagram⟩}\ \ -\omega^{-2}\text{⟨diagram⟩}\big)
$$

as claimed, and the "tuck" map is

$$\begin{pmatrix} \boxed{\;} & 0 & 0 \\ 0 & \smile\!\!\frown & \frac{1}{2}\smile\!\!\frown \end{pmatrix} \circ \begin{pmatrix} 1 \\ -\varphi^{-1}\lambda \\ 0 \end{pmatrix} = \begin{pmatrix} \boxed{\;} \\ \smile\!\!\frown \circ \,\rightleftharpoons \end{pmatrix}. \qquad \square$$

**Proof of Proposition 2.3** Finally, we'll construct explicit chain maps for the third Reidemeister move.

The complex associated to the left side is
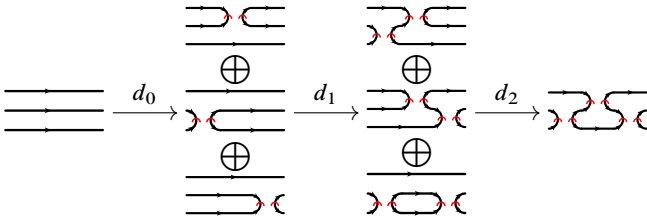


with differentials

$$d_0 = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix}$$

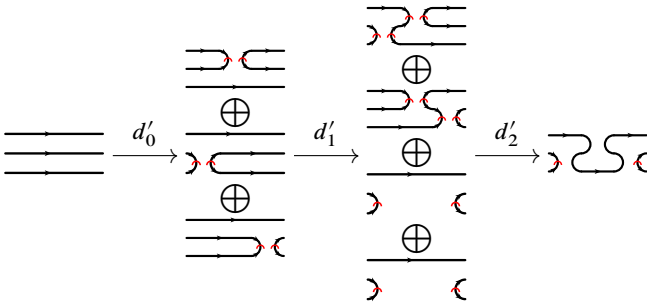$$d_1 = \begin{pmatrix} s_2 & -s_1 & 0 \\ s_3 & 0 & -s_1 \\ 0 & s_3 & -s_2 \end{pmatrix}$$

$$d_2 = \begin{pmatrix} s_3 & -s_2 & s_1 \end{pmatrix}.$$

We now need to simplify the complex; first delooping the last object at height two, and cancelling pairs of disorientations at height three using the isomorphisms

We obtain the complex



with differentials

$$d'_0 = d_0 = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix}$$



$$d'_2 = \zeta_{l3} d_2 \zeta_{l2}^{-1} = \boxed{\text{⟨diagram⟩}} \begin{pmatrix} s_3 & -s_2 & s_1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \text{⟨diagram⟩} & \frac{\omega^{-1}}{2}\,\text{⟨diagram⟩} \end{pmatrix}$$

$$= \left( \lambda = \left( \boxed{\text{⟨⟩}}\,s_3 \quad -\boxed{\text{⟨⟩}}\,s_2 \right) \quad \varphi = \omega^2 \mathbf{1} \; \bullet \right).$$

Applying the double Gaussian elimination lemma, we reach the homotopy equivalent complex

(A.2)

where

(A.3)
$$d_0'' = d_0' = \left(\;\vcenter{\hbox{}}\;\right)$$

(A.4)
$$d_1'' = \delta - \gamma\psi^{-1}\beta = \left(\;\vcenter{\hbox{}}\;\right)$$

via the simplifying (and unsimplifying) maps

$$\sigma_0 = \mathbf{1} \qquad\qquad\qquad \sigma_0^{-1} = \mathbf{1}$$

$$\sigma_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \qquad\qquad \sigma_1^{-1} = \left( \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -\psi^{-1}\beta \end{pmatrix} \right) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\sigma_2 = \left( \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \;\; -\gamma\psi^{-1} \right)\zeta_{l2} \qquad \sigma_2^{-1} = \zeta_{l2}^{-1} \begin{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ -\varphi^{-1}\lambda \\ \begin{pmatrix} 0 & 0 \end{pmatrix} \end{pmatrix}$$

$$\qquad = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -c_1 \end{pmatrix} \qquad\qquad\qquad = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -\omega^{-1}c_2 & \omega^{-1}c_3 \end{pmatrix}$$

$$\sigma_3 = 0 \qquad\qquad\qquad\qquad \sigma_3^{-1} = 0.$$

Here $c_1$ is the cobordism from  to  with three components, a disc, a curtain, and a saddle, $c_2$ the similar cobordism from  to  and $c_3$ is the similar cobordism from  to  (the adjoint of $c_1$).

That's half the work! Now we need to do the same for the right side of the third Reidemeister move, then compose a "simplifying map" with an "unsimplifying map".

Briefly, we calculate that the complex for the right side is

with differentials

$$d_0 = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix}$$

$$d_1 = \begin{pmatrix} s_2 & -s_1 & 0 \\ s_3 & 0 & -s_1 \\ 0 & s_3 & -s_2 \end{pmatrix}$$

$$d_2 = \begin{pmatrix} s_3 & -s_2 & s_1 \end{pmatrix}.$$

and, applying the simplification algorithm, that this is homotopy equivalent to the same complex as we obtained simplifying the other side of the Reidemeister move (shown in Equation (A.2)), but, somewhat tediously, with slightly different differentials

$$d_0'' = \begin{pmatrix} \phantom{xxx} \end{pmatrix}$$

$$d_1'' = \begin{pmatrix} \phantom{xxx} \end{pmatrix}.$$

These complexes thus differ by

$$\xi_0 = 1 \qquad \xi_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \xi_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The simplifying and unsimplifying maps are

$$\tau_0 = \mathbf{1} \qquad\qquad \tau_0^{-1} = \mathbf{1}$$

$$\tau_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \qquad\qquad \tau_1^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$$
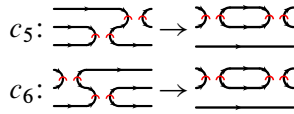
$$\tau_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -c_4 & 1 \end{pmatrix} \qquad\qquad \tau_2^{-1} = \begin{pmatrix} 1 & 0 \\ \omega^2 c_5 & \omega^2 c_6 \\ 0 & 1 \end{pmatrix}$$

$$\tau_3 = 0 \qquad\qquad \tau_3^{-1} = 0.$$

Here $\qquad\qquad c_4:$

$$c_5: \quad \longrightarrow$$

and

$$c_6: \quad \longrightarrow$$
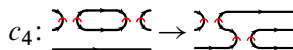
are the obvious variations on $c_1, c_2$ and $c_3$.

The interesting compositions, which provide us with the chain map between the two sides of the Reidemeister move, are

$$\tau_0^{-1} \circ \sigma_0 = (1)$$

$$\tau_1^{-1} \circ \sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \circ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

$$\tau_2^{-1} \circ \xi_2 \circ \sigma_2 = \begin{pmatrix} 1 & 0 \\ \omega^2 c_5 & \omega^2 c_6 \\ 0 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \circ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -c_1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ \omega^2 c_5 & -\omega^2 c_6 & \omega^2 c_6 c_1 \\ 0 & -1 & c_1 \end{pmatrix}.$$

The cobordism $c_6 c_1$ is the same "monkey saddle" appearing in [2].

The maps described in Proposition 2.3 describing the R3 chain map are simply a rearrangement of those presented here via matrices. $\qquad\square$

## A.3 Proofs of the R3 variations lemmas

We now turn to the proofs of Lemmas 2.7, 2.8 and 2.9. As explained previously, in Section 2.3.4, our strategy is to use the fact the Lemmas 2.4, 2.5 and 2.6 are exactly the special case that the Reidemeister 3 move is R3$_{\text{hml}}$, and then to show that if two Reidemeister 3 variations are adjacent in the cube of variations shown in Figure 10, and the spanning tree of definitions includes the connecting edge, then if the Lemmas hold for one variation, they must hold for the other. However, this approach immediately requires two cases, depending on whether we are looking at one of the four "vertical" edges of the cube of R3 variations, or one of the eight "horizontal edges".

The case that the connecting edge is "vertical", the formula defining one R3 variation in terms of the other (look back at Equation (2-3), for example, relating $R3_{\circlearrowright}$ and $R3_{hlm}$) involves conjugation by an R2 move in the direction opposite the highest crossing. On the other hand, when we look at the eight "horizontal" edges, the R2 move conjugation takes place opposite either the middle or lowest crossing. Because all of our lemmas are written describing the R3 moves in terms of resolutions of the highest crossing, it's unsurprising we need to treat these cases separately.

It turns out that in order to prove Lemma 2.8 for a given vertex $\star$, connected in the spanning tree to a vertex $\star'$, we'll have to know slightly more about the R3 map for $\star'$ than is explicit in the Lemmas. This extra information follows from the Lemmas however, and so we'll state it in the Corollary below. Once we have established the Lemmas for the vertex $\star'$ (starting at $\star' = $ hml), we also know the Corollary for $\star'$, and can use it in proving the Lemmas for $\star$.

**Corollary A.5** (Corollary of Lemma 2.7 and Lemma 2.8) *In the $\mathcal{P}$ layers of the cube of resolutions of $R3_\star$ there is exactly one resolution which appears for both the initial and final tangles of the R3 move. By grading considerations, the component of the R3 map between these resolutions is some multiple of a map whose underlying unoriented surface is the identity. There is always a unique configuration of seams on this surface without loops, and we will write "$\mathbf{1}$" for such a "disorientation cylinder with minimal seams". Write $p_\star$ for the coefficient of this disoriented surface. Writing $\kappa_\star$ for the coefficient appearing in the lowest homological height of the $\mathcal{O} \to \mathcal{O}$ map, and $\lambda_\star$ for the coefficient in the highest height (so by Lemma 2.8, $\lambda_\star = \kappa_\star$ if $\star = $ hml, lhm, mhl or lmh, and $\lambda_\star = -\omega^2 \kappa_\star$ otherwise), we have*

$$\frac{p_\star}{\kappa_\star} = \begin{cases} -1 & \text{if } \star = \text{hml, lmh, mlh or } \circlearrowright, \\ \omega^2 & \text{if } \star = \text{lhm, mhl, hlm or } \circlearrowleft \end{cases}$$

**Proof** This is actually quite involved! Along the way, we'll also need to understand one of the coefficients in the "downhill" map. We'll introduce some further notation for particular resolutions of the R3 tangle, as follows: a symbol $abc$, with each of $a$ and $b$ either $>$ or $<$, and $c$ either $\mathcal{O}$ or $\mathcal{P}$, refers to the resolution in which the first crossing is either in the higher or lower homological height resolution, depending on $a$, the second crossing is again either in the higher or lower homological height resolution, depending on $b$, and the third crossing is either in the orthogonal or parallel resolution relative to the triangle formed by the R3 tangle, depending on $c$. Remember that the convention for the ordering of crossings is unobvious; before an R3 move (when the triangle is on the left of the lowest strand), the crossings are ordered as "middle" then

"low" then "high", while after the R3 move the crossings are ordered as "low" then "middle" then "high".

It's easy to verify that in the $\mathcal{O} \to \mathcal{P}$ cases, the $><\mathcal{O}$ resolution of the initial tangle is the same, ignoring orientation data, as the $<<\mathcal{P}$ resolution of the final tangle. Since these resolutions are in the same $q$–grading, the only maps between them are disoriented cylinders. Taking into account orientation data, we claim that there is a unique allowed configuration of disorientation seams on the cylinder with the appropriate boundary. We then define $q_\star$ to be the coefficient appearing on this map in the R3$_\star$ map. Similarly, in the $\mathcal{P} \to \mathcal{O}$ cases, the $>>\mathcal{P}$ resolution of the initial tangle is the same up to orientation data as the $><\mathcal{O}$ resolution of the final tangle, and we define $q_\star$ to be the coefficient appearing on coefficient of the component of the R3 map between these resolutions.

The resolutions described in the statement of the corollary are in this notation $><\mathcal{P}$ (in the initial tangle) and $<>\mathcal{P}$ (in the final tangle), in the $\mathcal{O} \to \mathcal{P}$ cases, and the reverse in $\mathcal{P} \to \mathcal{O}$ cases.

We now determine $q_\star$, and then $p_\star$ in terms of $\kappa_\star$, by considering the following two pairs of commuting squares coming from chain map conditions.

(A.5)
$$
\begin{array}{ccc}
<<\mathcal{O} & \xrightarrow{\;\kappa_\star \mathbf{1}\;} & <<\mathcal{O} \\
{\scriptstyle (-1)^{\star=\text{lhm or mlh}}s}\downarrow & & \downarrow{\scriptstyle +s} \\
><\mathcal{O} & \xrightarrow{\;q_\star\text{“}\mathbf{1}\text{”}\;} & <<\mathcal{P} \\
{\scriptstyle +s}\downarrow & & \downarrow{\scriptstyle (-1)^{\star=\text{hml or lhm}}s} \\
><\mathcal{P} & \xrightarrow{\;p_\star\text{“}\mathbf{1}\text{”}\;} & <>\mathcal{P}
\end{array}
$$

(A.6)
$$
\begin{array}{ccc}
>>\mathcal{O} & \xrightarrow{\;\lambda_\star \mathbf{1}\;} & >>\mathcal{O} \\
{\scriptstyle +s}\uparrow & & \uparrow{\scriptstyle (-1)^{\star=\text{hlm or }\cup}s} \\
>>\mathcal{P} & \xrightarrow{\;q_\star\text{“}\mathbf{1}\text{”}\;} & ><\mathcal{O} \\
{\scriptstyle (-1)^{\star=\cup\text{ or lmh}}s}\uparrow & & \uparrow{\scriptstyle +s} \\
<>\mathcal{P} & \xrightarrow{\;p_\star\text{“}\mathbf{1}\text{”}\;} & ><\mathcal{P}
\end{array}
$$

The signs appearing on saddles in Equations (A.5) and (A.6) are calculated by the usual rule for sprinkling signs in tensor products of complexes (see Section A.6.1), and the convention for ordering crossings before and after Reidemeister 3 moves.

These calculations tell us that

$$
p_\star = \begin{cases} -\kappa_\star & \text{if } \star = \text{hml}, \text{mlh}, \text{lmh or } \circlearrowright, \\ \omega^2 \kappa_\star & \text{if } \star = \text{hlm}, \text{lhm}, \text{mhl or } \circlearrowleft \end{cases}
$$

$$
= \begin{cases} -1 & \text{if } \star = \text{hml or lmh}, \\ 1 & \text{if } \star = \text{hlm or mlh}, \\ \omega^2 & \text{if } \star = \text{lhm or mhl}, \\ -\omega^2 & \text{if } \star = \circlearrowright \text{ or } \circlearrowleft. \end{cases}
$$

Unfortunately there's a small subtlety in extracting the relation between $k_\star$ and $p_\star$ from Equations (A.5) and (A.6); the horizontal arrows are not labelled by multiples of the identity map, but by multiples of the "minimal seam" identity map. Depending on the configuration of these seams, it might not be the case that $s\text{"}\mathbf{1}\text{"} = \text{"}\mathbf{1}\text{"}s$, but that they differ by a power of $\omega$. This requires a case by case analysis. Defining $\sigma_\star$ and $\tau_\star$ so that $s\text{"}\mathbf{1}\text{"} = \sigma_\star\text{"}\mathbf{1}\text{"}s$ in the upper squares of Equations (A.5) and (A.6), and $s\text{"}\mathbf{1}\text{"} = \tau_\star\text{"}\mathbf{1}\text{"}s$ in the lower squares, we find that all $\sigma_\star$ and $\tau_\star$ are equal to 1, except that

$$
\sigma_{\text{lhm}} = \sigma_{\text{mhl}} = \tau_{\circlearrowright} = \tau_{\circlearrowleft} = \omega^2.
$$

The corollary now follows. □

Let's begin the "vertical" edge case by introducing some notation for particular subspaces of the complexes associated to the four tangles appearing in our formula for one R3 move in terms of another. The symbols $\mathcal{O}|$ and $\mathcal{P}|$ will denote the spaces of the Khovanov complex of the initial tangle in which the highest crossing has been resolved in the orthogonal and parallel manners respectively. The symbols $|\mathcal{O}$ and $|\mathcal{P}$ will denote the corresponding subspaces of the final tangle. The symbols $a|bc$ and $ab|c$, where $a, b, c = \mathcal{O}$ or $\mathcal{P}$ will denote subspaces of the two intermediate tangles (the lower left and lower right tangles in Equation (2-3), respectively), in which the three crossings not involving the lowest strand (that is, the original highest crossing, and the two new crossings introduced by the R2 move) have been resolved either orthogonal or parallel to the lowest strand, according to the values of $a, b$ and $c$, with $a$ referring to the original highest crossing, $b$ referring to the new crossing closest to the original tangle, and $c$ to the new crossing furthest away. (The vertical bar | is meant to denote the lowest strand.)

We know, from Figure 7 and Figure 8, that the R2 maps only see those subspaces in which the two crossings involved in the R2 move have been resolved the same way, that is, with $b = c$. Thus if R3$_{\star'}$ is defined in terms of R3$_\star$ as the composition of an

R2 map, the map $R3_\star^{-1}$, and an inverse R2 map, as in the example in Equation (2-3), it will have the form shown in Figure 19.
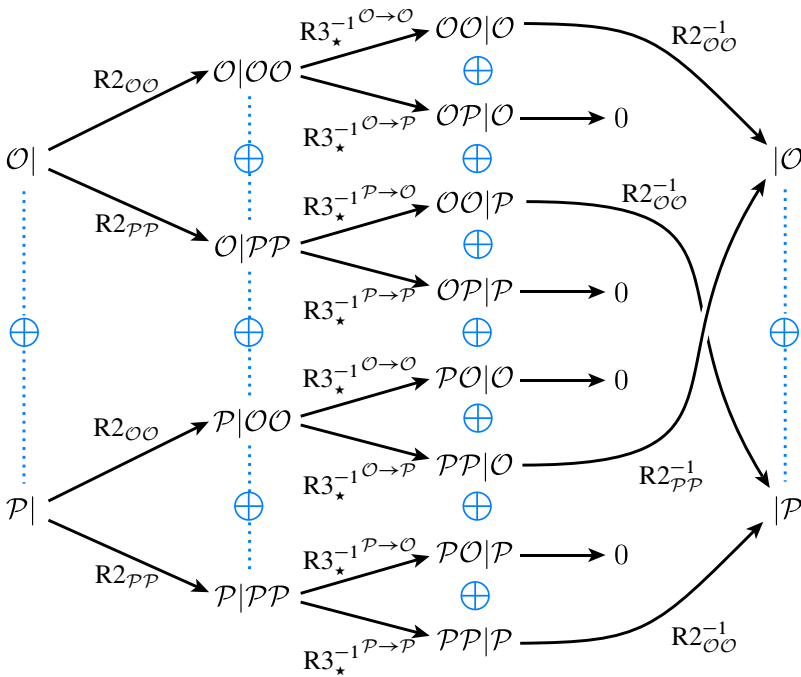


Figure 19: One R3 variation defined in terms of another via R2 moves

**Proof of Lemma 2.7** (Vertical edge cases)  If the statement of the Lemma holds for some move $R3_\star$, and we're defining another $R3_{\star'}$ in terms of it via a vertical edge of the cube in Figure 10, it must also hold for $R3'_\star$. (Recall that adjacent R3 moves in the cube have opposite arrangements of layers.) This is the case simply because in Figure 19 the map $R3_{\star'}^{\mathcal{P}\to\mathcal{O}}$ factors through $R3_\star^{-1\mathcal{O}\to\mathcal{P}}$, and the map $R3_{\star'}^{\mathcal{O}\to\mathcal{P}}$ factors through $R3_\star^{-1\mathcal{P}\to\mathcal{O}}$. $\qquad\square$

**Proof of Lemma 2.8** (Vertical edge cases)  Using Lemma 2.7, one sees that the component $R3_\star^{\mathcal{O}\to\mathcal{O}}$ is itself a chain map. This follows in the case that the layers are arranged as $\mathcal{O}\to\mathcal{P}$ by writing $d = d_{\mathcal{O}\to\mathcal{O}} + d_{\mathcal{O}\to\mathcal{P}} + d_{\mathcal{P}\to\mathcal{P}}$; then the $\mathcal{O}\to\mathcal{O}$ component of the equation $dR3_\star = R3_\star d$ simply says

$$d_{\mathcal{O}\to\mathcal{O}}R3_\star^{\mathcal{O}\to\mathcal{O}} = R3_\star^{\mathcal{O}\to\mathcal{O}}d_{\mathcal{O}\to\mathcal{O}} + R3_\star^{\mathcal{O}\to\mathcal{P}}d_{\mathcal{P}\to\mathcal{O}}$$

and since by Lemma 2.7 $R3_\star^{\mathcal{P}\to\mathcal{O}} = 0$,

$$d_{\mathcal{O}\to\mathcal{O}}R3_\star^{\mathcal{O}\to\mathcal{O}} = R3_\star^{\mathcal{O}\to\mathcal{O}}d_{\mathcal{O}\to\mathcal{O}}.$$

The other case, in which the layers are arranged as $\mathcal{P} \to \mathcal{O}$, is the same.

We claim then that the only chain maps from one orthogonal layer to another are multiples of the identity in the $\star = \mathrm{hml}, \mathrm{lhm}, \mathrm{mhl}$ or $\mathrm{lmh}$ cases, and multiples of the standard chain map described in the Lemma when $\star = \mathrm{hlm}, \mathrm{mlh}, \circlearrowleft$ or $\circlearrowright$. The overall coefficient is easily determined from Figure 19, and the formulas for the R2 maps in Figure 7 and Figure 8. We obtain the result described in the Lemma, that for $\star = \mathrm{hml}, \mathrm{lhm}, \mathrm{mhl}$ or $\mathrm{lmh}$ the component in lowest homological height is actually the identity, that for $\star = \mathrm{hlm}$ and $\circlearrowright$ the coefficient of the component in the lowest homological height is $\omega^2$, and that for $\star = \mathrm{mlh}$ and $\circlearrowleft$ that coefficient is $-1$.  □

**Proof of Lemma 2.9** (Vertical edge cases)  Again looking at Figure 19, we see that $R3_{\star'}^{\mathcal{P}\to\mathcal{P}}$ factors through $R3_\star^{-1\mathcal{P}\to\mathcal{P}}$. Thus if the lemma holds for $R3_\star$ (which it does for $\star = \mathrm{hml}$, by Lemma 2.6), it also holds for any adjacent R3 move which we're defining in terms $R3_\star$. The second part of the Lemma, describing the normalisation, has already been proved as part of Corollary A.5. The final statement, about the other entries of the map in the middle homological height having disc components, follows immediately from grading considerations.  □

We now deal with the cases involving a horizontal edge.

**Proof of Lemma 2.7** (Horizontal edge cases)  We'll introduce a new notational convention; when decorating a crossing with an $\mathcal{O}$ or a $\mathcal{P}$, to indicate a particular resolution, we'll also draw a short squiggly line pointing towards the nearby strand with respect to which we mean "orthogonal" or "parallel". We can rotate this short squiggly line into a different region adjacent to the crossing, if at the same time we interchange the labels $\mathcal{O}$ and $\mathcal{P}$.

When we define an R3 map via a "horizontal" edge in the cube, in terms of some other map, say $R3_\star$, it has the form:



(A.7)

(Here the labelled crossings in the initial and final step are the highest crossings, as usual.) Thus we see that the $\mathcal{O} \to \mathcal{P}$ component factors through the original $\mathcal{P} \to \mathcal{O}$ component of R3$_\star$, and similarly the $\mathcal{P} \to \mathcal{O}$ component factors through the $\mathcal{O} \to \mathcal{P}$ component. Since the relative heights of the $\mathcal{O}$ and $\mathcal{P}$ maps are reversed in adjacent R3 variations in the cube, this suffices to establish the lemma. □

**Proof of Lemma 2.8** (Horizontal edge cases) This follows the argument above in the vertical edge case; Lemma 2.7, ensures that the component R3$_\star^{\mathcal{O} \to \mathcal{O}}$ is a chain map, and thus only multiples of the map described in this Lemma are possible. To check that the multiple is the one described, we follow through the $\mathcal{O} \to \mathcal{O}$ composition in Equation (A.7) above. Notice that this relies on Corollary A.5, for the normalisation of the $\mathcal{P} \to \mathcal{P}$ map appearing in Equation (A.7). Further, in the cases where the R2 moves appearing are R2b moves, one must take into account a sign of homological origin, coming from reordering crossings. □

**Proof of Lemma 2.9** (Horizontal edge cases) We now look in slightly more detail at Equation (A.7). The highest and lowest homological heights of the $\mathcal{P}$ layer consist of those resolutions in which the other two crossings (ie, the middle and lowest crossing) have been resolved in opposite ways; one as a $\mathcal{O}$, one as a $\mathcal{P}$. We look at one of the two cases, the other being essentially identical. Making use of Lemma 2.8 (in particular, that, ignoring all disorientation data and coefficients, the $\mathcal{O} \to \mathcal{O}$ components of all R3 variations are simply the identity), we see

(A.8)
$$
\text{R2} \quad = \quad \text{R3}_\star^{\mathcal{O} \to \mathcal{O}} \quad \text{R2}^{-1} \quad 0 .
$$

The second part of the Lemma, describing the normalisation, has already been proved as part of Corollary A.5. The final statement, about the other entries of the map in the middle homological height having disc components, follows immediately from grading considerations. □

## A.4 Planar algebras and canopolises

A planar algebra is a gadget specifying how to combine objects in planar ways. They were introduced by Jones [11] to study subfactors, and have since found more general use.

In the simplest version, a planar algebra $\mathcal{P}$ associates a vector space $\mathcal{P}_k$ to each natural number $k$ (thought of as a disc in the plane with $k$ marked points on its boundary)

and a linear map $\mathcal{P}(T)\colon \mathcal{P}_{k_1} \otimes \mathcal{P}_{k_2} \otimes \cdots \otimes \mathcal{P}_{k_r} \to \mathcal{P}_{k_0}$ to each planar tangle[4] $T$, for example
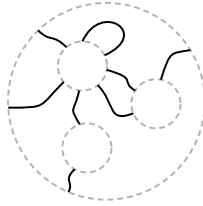


with internal discs with $k_1, k_2, \ldots, k_r$ marked points, and $k_0$ marked points on the external disc. These maps (the "planar operations") must satisfy certain properties: "radial" tangles induce identity maps, and composition of the maps $\mathcal{P}(T)$ is compatible with the obvious composition of planar diagrams by gluing one inside the other.

For the exact details, which are somewhat technical, see Jones [11].

Planar algebras also come in more subtle flavors. Firstly, we can introduce a label set $\mathfrak{L}$, and associate a vector space to each disc with boundary points marked by this label set. (The simplest version discussed above thus has a singleton label set, and the discs are indexed by the number of boundary points.) The planar tangles must now have arcs labeled using the label set, and the rules for composition of diagrams require that labels match up. Secondly, we needn't have vector spaces and linear maps between them; a planar algebra can be defined over an arbitrary monoidal category, associating objects to discs, and morphisms to planar tangles. Thus we might say "$\mathcal{P}$ is a planar algebra over the category $\mathcal{C}$ with label set $\mathfrak{L}$." [5]

A "canopolis", introduced by Bar-Natan in [2][6][7], is simply a planar algebra defined over some category of categories, with monoidal structure given by cartesian product. Thus to each disc, we associate some category of a specified type. A planar tangle then induces a functor from the product of internal disc categories to the outer disc category, thus taking a tuple of internal disc objects to an external disc object, and a tuple of internal disc morphisms to an external disc morphism. It is picturesque to think of the objects living on discs, and the morphisms in "cans", whose bottom and top surfaces correspond to the source and target objects. Composition of morphisms is achieved by stacking cans vertically, and the planar operations put cans side by side.

---

[4]Familiarly known as a "spaghetti and meatballs" diagram.

[5]A "subfactor planar algebra" is defined over Vect, and has a 2 element label set. We impose an additional condition that only discs with an even number of boundary points and with alternating labels have nontrivial vector spaces attached. There is also a positivity condition. See Bar-Natan [3, Section 4].

[6]He called it a "canopoly", instead, but we're taking the liberty of fixing the name.

[7]See also Webster [22] for a description of Khovanov–Rozansky homology [16; 17] using canopolises.

The functoriality of the planar algebra operations ensure that we can build a "city of cans" (hence the name canopolis) any way we like, obtaining the same result: either constructing several "towers of cans" by composing morphisms, then combining them horizontally, or constructing each layer by combining the levels of all the towers using the planar operations, and then stacking the levels vertically.

## A.5  Complexes in a canopolis form a planar algebra

Given a quadratic tangle,
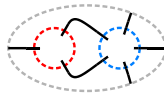


and a pair of complexes associated to the inner discs,



we need to define a new complex associated to the outer disc.

We'll imitate the usual construction for tensor product of complexes, but use the quadratic tangle to combine objects and morphisms. Form a double complex then collapse along the antidiagonal:
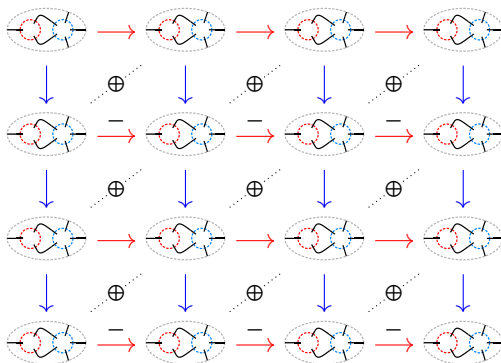


Here each horizontal arrow is the planar composition of a morphism from $C_1$, placed in the left disc, with the identity on the appropriate object from $C_2$, in the right disc. Similarly, each vertical arrow is the planar composition of a morphism from $C_2$ with an identity morphism.

The extension to tangles with more than 2 internal discs is obvious. Moreover, it's not hard to see that chain maps between complexes in a canopolis also form a planar algebra, providing the morphism part of "the canopolis of complexes and chain maps".

## A.6  Homological conventions

**A.6.1  Tensor product**  In the next two sections we'll describe certain conventions to do with tensoring complexes. (Please accept our apologies if they're not what you're used to!) See Gelfand and Manin [9].

The tensor product of two complexes $(A^\bullet, d_A)$ and $(B^\bullet, d_B)$ is defined to be

$$(A \otimes B)^\bullet = \bigoplus_{i+j=\bullet} A^i \otimes B^j,$$

$$d_{(A \otimes B)^\bullet} = \sum_{i+j=\bullet} (-1)^j d_{A^i} \otimes \mathbf{1}_{B^j} + \mathbf{1}_{A^i} \otimes d_{B^j}.$$

If you think of $A^\bullet$ as lying horizontally, and $B^\bullet$ as vertically, this rule says "negate the differentials in every odd row".

**A.6.2  Permuting tensor products**  Unfortunately, while $A^\bullet \otimes B^\bullet \cong B^\bullet \otimes A^\bullet$ the isomorphism can't just be the identity. Instead, we'll take it to be $A^i \otimes B^j \mapsto (-1)^{ij} B^j \otimes A^i$; that is it negates anything in "doubly odd" degree.

The only complexes we ever take tensor products of are the complexes associated to tangles. In the simplest case, where we are taking the tensor product of two crossings, the "crossing reordering" map is "negate doubly disoriented smoothings". That is, objects in which both crossings have been resolved in the disoriented direction get negated when we change the ordering of the crossings.

# References

[1]  **D Bar-Natan**, *Khovanov homology for knots and links with up to 11 crossings*, from: "Advances in topological quantum field theory", (J M Bryden, editor), NATO Sci. Ser. II Math. Phys. Chem. 179, Kluwer Acad. Publ., Dordrecht (2004) 167–241  MR2147420

[2]  **D Bar-Natan**, *Khovanov's homology for tangles and cobordisms*, Geom. Topol. 9 (2005) 1443–1499  MR2174270

[3]  **D Bar-Natan**, *Fast Khovanov homology computations*, J. Knot Theory Ramifications 16 (2007) 243–255  MR2320156

[4]  **J W Barrett**, **B W Westbury**, *Spherical categories*, Adv. Math. 143 (1999) 357–375  MR1686423

[5] **D Bisch**, *Bimodules, higher relative commutants and the fusion algebra associated to a subfactor*, from: "Operator algebras and their applications (Waterloo, ON, 1994/1995)", (P A Fillmore, J A Mingo, editors), Fields Inst. Commun. 13, Amer. Math. Soc. (1997) 13–63 MR1424954

[6] **C L Caprau**, *sl(2) tangle homology with a parameter and singular cobordisms*, Algebr. Geom. Topol. 8 (2008) 729–756 MR2443094

[7] **J S Carter**, **J H Rieger**, **M Saito**, *A combinatorial description of knotted surfaces and their isotopies*, Adv. Math. 127 (1997) 1–51 MR1445361

[8] **J S Carter**, **M Saito**, *Reidemeister moves for surface isotopies and their interpretation as moves to movies*, J. Knot Theory Ramifications 2 (1993) 251–284 MR1238875

[9] **S I Gelfand**, **Y I Manin**, *Methods of homological algebra*, Springer, Berlin (1996) MR1438306 Translated from the 1988 Russian original

[10] **M Jacobsson**, *An invariant of link cobordisms from Khovanov homology*, Algebr. Geom. Topol. 4 (2004) 1211–1251 MR2113903

[11] **V F R Jones**, *Planar algebras, I* arXiv:math.QA/9909027

[12] **A Joyal**, **R Street**, *The geometry of tensor calculus. I*, Adv. Math. 88 (1991) 55–112 MR1113284

[13] **M Khovanov**, *A categorification of the Jones polynomial*, Duke Math. J. 101 (2000) 359–426 MR1740682

[14] **M Khovanov**, *A functor-valued invariant of tangles*, Algebr. Geom. Topol. 2 (2002) 665–741 MR1928174

[15] **M Khovanov**, *Categorifications of the colored Jones polynomial*, J. Knot Theory Ramifications 14 (2005) 111–130 MR2124557

[16] **M Khovanov**, **L Rozansky**, *Matrix factorizations and link homology*, Fund. Math. 199 (2008) 1–91 MR2391017

[17] **M Khovanov**, **L Rozansky**, *Matrix factorizations and link homology. II*, Geom. Topol. 12 (2008) 1387–1425 MR2421131

[18] **R Kirby**, **P Melvin**, *The 3–manifold invariants of Witten and Reshetikhin–Turaev for* sl(2, **C**), Invent. Math. 105 (1991) 473–545 MR1117149

[19] **E S Lee**, *An endomorphism of the Khovanov invariant*, Adv. Math. 197 (2005) 554–586 MR2173845

[20] **G Naot**, *The universal Khovanov link homology theory*, Algebr. Geom. Topol. 6 (2006) 1863–1892 MR2263052

[21] **D Roseman**, *Reidemeister-type moves for surfaces in four-dimensional space*, from: "Knot theory (Warsaw, 1995)", (V F R Jones, P T Joanna Kania-Bartoszyńska, Jozef H Przytycki, V G Turaev, editors), Banach Center Publ. 42, Polish Acad. Sci., Warsaw (1998) 347–380 MR1634466

[22]  **B Webster**, *Khovanov–Rozansky homology via a canopolis formalism*, Algebr. Geom. Topol. 7 (2007) 673–699  MR2308960

DC:  *Department of Mathematics, Randolph-Macon College*
*Ashland, VA 23005, USA*

SM,KW:  *Microsoft Station Q, University of California*
*Santa Barbara 93106-6105, USA*

davidclark@rmc.edu,  scott@tqft.net,  kevin@canyon23.net

http://faculty.rmc.edu/davidclark,  http://tqft.net/,
http://canyon23.net/math/