Properties of Knotoids and $\mathbb{Z}$-Tangles

by

Larsen Drew Linov

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Mathematics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Ian Agol, Chair
Professor John Lott
Professor Alistair Sinclair

Spring 2024

Properties of Knotoids and $\mathbb{Z}$-Tangles

Abstract

Properties of Knotoids and $\mathbb{Z}$-Tangles

by

Larsen Drew Linov

Doctor of Philosophy in Mathematics

University of California, Berkeley

Professor Ian Agol, Chair

In addition to classical knot theory, low-dimensional topology is also concerned with a variety of alternative classes of knot-like objects, including knotoids, braids, and string links. In this thesis we explore several of these theories as well as certain connections between them. While our results cover several different types of structures, they nonetheless build on each other.

The theory of knotoids is an extension of classical knot theory whereby knot diagrams are allowed to have two endpoints instead of being a closed loop. One of the most important knotoid invariants is the height h, which measures how far a knotoid is from being a classical knot. After defining the signed versions h_+ and h_- of the height, we prove that together they determine the unsigned height by the simple formula $h_+ + h_- = h$, and we demonstrate a few applications.

String links, which are tangles that have been normalized with respect to the locations of the endpoints of the strands, come with a natural monoid operation that generalizes both connected summation of knots and composition of braids. Previous work on strings links has shown that the string link monoids are cancellative up to multiplication with units. Here we strengthen that result by proving the freeness of unit multiplication on either side, thus answering a question of Blair–Burke–Koytcheff and implying full cancellativity of the string link monoids. In the same section, we prove a weaker version of the well-known cosmetic crossing conjecture and explore applications of knotoids to braid groups.

We also study an infinite version of string links, which we call $\mathbb{Z}$-tangles. These $\mathbb{Z}$-tangles are interesting in part because a version of Birman–Hilden theory applies to $\mathbb{Z}$-tangles as they relate to knotoids. Specifically, there is a natural forgetful function from the set of knotoids to the set of $\mathbb{Z}$-tangles up to isotopy. We give several pieces of evidence for the conjectural injectivity of this Birman–Hilden map, most notably by using some of the results of earlier chapters to identify classes of knotoids that will each be uniquely determined by their image.

Contents

Acknowledgments

I am very lucky to have been supported, both academically and personally, by so many excellent people during my time in grad school. My years in Berkeley have been incredibly rewarding, and it has been a group effort the entire way.

First I'd like to thank my advisor, Ian Agol, for his deep knowledge and generous patience.

I'm thankful also to Mom, Dad, Shayna, and my broader family for always believing in me and always being there for me.

I am indebted to Roy Zhao, James Dix, James Rowan, and every other member of my awesome cohort for all that they've done to lift me up over the years.

Thanks also to my fellow topologists at Berkeley, to my sports teammates, to my officemates, my MGSA collaborators, the math department staff, and many others who have given me comfort, guidance, and purpose.

Chapter 1

Introduction

1.1 Background and Context

Knot theory has been an area of earnest study since even before topology emerged as a recognized subfield of math, with the Tait conjectures ([51]) and the Alexander polynomial ([3]) marking advancements in an increasingly formalized theory of knots. Originally inspired by artwork, physics, and of course the practical uses of tying ropes in daily life, the mathematical idea of a knot as a closed loop in three-dimensional space has in turn generated various offshoots and generalizations, among them tangles, braids ([5]), spatial graphs ([15]), and virtual knots ([29]).

One such type of knot-like object, dating to 1990, is a string link. String links, which are tangles for which the ambient 3-ball has a designated "top" and "bottom" and for which each strand has one endpoint on each side, were defined in [26] (Definition 1.1) by Habegger and Lin in order to tackle Milnor's problem of classifying links up to the equivalence relation of link-homotopy ([41], page 194). (We will define string links more formally in Chapter 2.) The utility of string links for studying classical links is due to the fact that that the set of string links (with a fixed number of strings) forms a monoid with stacking as the binary operation. Since string links capture a lot of general behavior in knot theory, recent efforts have explored the properties of the string link monoids, such as primality, centrality, and cancellativity ([13, 8, 48]).

More recently, in [49] (Section 2.1), Turaev introduced knotoids in order to better compute the Seifert genera and knot groups of classical knots. A knotoid is, in brief, a path between two points on a sphere, with extra restrictions and information where the path self-intersects, considered up to ambient isotopy and Reidemeister moves. Like virtual knots ([22]), knotoids can be considered to be an extension of standard knot theory in that there is a natural injection from the set of (oriented) knot classes to the set of knotoids. They are closely related to spatial graphs ([49], Section 5.1) and strongly invertible knots ([7], Theorem 1.1) in addition to finding applications in biology (see, e.g., [6, 21]). Much of the work previously done on knotoids relates to the height invariant (e.g. [23, 33]), which is a geometric measure of how far a knotoid is from being a classical knot.

Meanwhile, in the broader area of geometric topology, one of the other major topics of interest besides knot theory is the theory of mapping class groups of surfaces. (Given a topological space, the associated mapping class group is the group of connected components of the group of automorphisms of the space, if the latter set is given an appropriate topology—see [19], Section 2.1.) In a series of papers in the 1970s ([11, 9, 10], see also [39]), Birman and Hilden investigated the relationship between the mapping class groups of two surfaces when one covers the other. Specifically, in order to ease the computation of one mapping class group after already knowing the other, they unpacked a sequence of intermediate group homomorphisms and found sufficient conditions for the most difficult intermediate map to be injective. While we will not actually use classical Birman–Hilden theory in this thesis, we will define a knotoid invariant analogous to the aforementioned map and call it the "Birman–Hilden map."

We also note a general theme in geometric topology of generalizing problems to non-compact contexts. A salient example of this is the very active subject of "big mapping class groups", that is, mapping class groups of surfaces whose fundamental groups are not finitely generated ([4]). While research on big mapping class groups usually disregards surfaces with non-compact boundary, such surfaces are nonetheless still occasionally studied ([17]), for example in the version of the infinite-stranded braid group defined by [18]. Going back to three dimensions, in the realm of knot theory there is also interest in links of infinitely many components, especially those coming as preimages of links in nontrivial 3-manifolds (see, e.g., [16, 47]).

1.2 Overview

This thesis is organized as follows:

The present chapter is a minimal introduction, deferring much of the preliminary material to Chapter 2. However, Section 1.3 is included here to preemptively clarify potential sources of ambiguity.

Chapter 2 lays the groundwork for the rest of the document, containing a relatively thorough account of the objects and operations that will appear later on. It includes most of the formal definitions that we will use, including definitions for knotoids, $\mathbb{Z}$-tangles and annular 1-tangles, string links, $\mathbb{Z}$-braids and $\mathbb{N}$-braids, height, and the Birman–Hilden map for knotoids. At the same time, Chapter 2 also touches on basic facts about these objects, providing proofs where appropriate. The later chapters also include a few definitions of their own, but only when the concept in question will appear only briefly.

The bulk of our results are contained in Chapters 3 through 5.

Chapter 3, which roughly matches the material previously published as [36], focuses on signed heights of knotoids and annular 1-tangles. It states and proves the additive relationship between signed and unsigned heights, and it gives simple lower bounds on the signed heights, which in turn can improve lower bounds on the unsigned height of a knotoid. Finally, this chapter includes a categorization, and some degree of classification, of knotoids with height up to 2.

Chapter 4 is more eclectic, with some parts that aren't strictly related to the others, but the chapter does follow a somewhat consistent theme. It starts by identifying a family of string link invariants that, taken together, distinguish distinct units in the string link monoids. This is used to address a question appearing in [12]. A special case of that problem can be stated in terms of group actions by mapping class groups on certain sets of isotopy classes of tangles; as a tangent, we also address a slightly different version of that problem in which the tangles are not string links but other types of 2-tangles. The latter problem is related to the cosmetic crossing conjecture, but there is no obvious way to use our results to prove the standard version of the conjecture. Next, Section 4.2 proves a pair of similar facts about self-similarity in $\mathbb{Z}$-braids and $\mathbb{N}$-braids, respectively. The first is proved using basic hyperbolic geometry and is used as a lemma later on; in contrast, the statement for $\mathbb{N}$-braids does not get applied in this thesis but is included as an example of an external application of knotoids and the Birman–Hilden map. In the last section of Chapter 4, some of the lemmas from the previous sections are used to prove a geometric fact about the monoid structure of a subset of $\mathbb{Z}$-tangles.

Chapter 5 mainly contains a collection of partial results backing up the conjecture that the Birman–Hilden map is injective. In particular, most of the work in this chapter goes toward finding restrictions on the heights and signed heights of possible counterexamples to injectivity. To this end we rely heavily on the results of Section 4.1 and the strategies in Section 2 of [44].

Finally, Chapter 6 acts as something of an appendix, containing tangential information that has been moved away from the main chapters to maintain a more even flow. It can also be considered as a sister chapter to Chapter 5, since that is the chapter to which it is most related. The ideas in Chapter 6 have not been studied to their limits, so it includes questions available as directions for future research.

1.3 Conventions and Terminology

Throughout this document, manifolds will be oriented and smooth, often with boundary and corners. Maps between manifolds will usually be smooth, or sometimes piecewise-smooth. In our contexts, corners and piecewise-smooth maps between manifolds will typically be smoothable in a canonical-enough way, so without causing any significant ambiguity we will take some liberty in treating the two conventions as interchangeable. For example, "diffeomorphisms" will sometimes not be smooth but merely piecewise-smooth; thus in the case of smooth manifolds with corners, diffeomorphisms will not generally be required to send corners to corners.

Some of the manifolds we will work with are non-compact manifolds with boundary. In this context there are two competing meanings of the word "proper", so to prevent ambiguity we will use the following terms instead: A map is *end-proper* if the preimage of any compact set is compact, and an embedding of manifolds $f\colon M \to M'$ is *∂-proper* if $f^{-1}(\partial M') = \partial M$ and $f(M)$ is transverse to $\partial M'$. If a ∂-proper embedding is also end-proper, we will call it *biproper*.

In an abuse of notation, we will frequently conflate an embedding map with its image when doing so would not create confusion.

When we write S^1 we will always specifically mean $\mathbb{R}/\mathbb{Z}$, as opposed to $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$ or any other manifestation of a circle. To discuss specific elements we will sometimes, for example, write "0" to refer to the the class containing 0.

The symbol I will always be the closed interval $[-1, 1]$. It should not be confused with $\mathcal{I}$, which will be defined later on.

The word "knot" will most often be used to mean an isotopy class of oriented knots in S^3, but we will not be strict in this convention. When we discuss a specific representative of a knot or an unoriented knot we will usually make it clear that we are doing so. We will treat isotopy classes of knots in S^3 as interchangeable with isotopy classes of 1-tangles in a 3-ball, according to the standard correspondence.

Chapter 2

The Objects in Question

In this chapter we will set the stage for the later chapters by laying out definitions for most of the objects, functions, operations, and invariants that will be under consideration.

2.1 $\mathbb{Z}$-Tangles

Recall that I denotes the interval $[-1, 1]$.

Definition 2.1.1. A $\mathbb{Z}$-*tangle* is a biproper embedding $\mathbb{Z} \times I \to \mathbb{R} \times I^2$ with $(n, \pm 1) \mapsto (n, \pm 1, 0)$ for all $n \in \mathbb{Z}$.

See Figure 2.1 for an example. We will denote the set of all $\mathbb{Z}$-tangles by A. Note that a $\mathbb{Z}$-tangle is like a tangle in the sense that $\mathbb{R} \times I^2$ is the same as a 3-ball with two boundary points removed. Because of our restrictions on the locations of the endpoints of each strand, it can be even better to think of $\mathbb{Z}$-tangles as string links (which will be defined later in this section) of infinite size.

When discussing $\mathbb{R} \times I^2$, we will refer to the three coordinates as x, y, and z, in that order, as is standard for subsets of $\mathbb{R}^3$.

Isotopy and Translations

We will use the language of group actions to discuss isotopy between $\mathbb{Z}$-tangles: Let $\mathcal{I}$ be the set of boundary-fixing automorphisms of $\mathbb{R} \times I^2$. There is a natural action of $\mathcal{I}$ on A by $(\iota, \ell) \mapsto \iota \circ \ell$. We claim that two $\mathbb{Z}$-tangles are in the same orbit if and only if they are related by boundary-fixing ambient isotopy. The boundary-fixing automorphism group of a 3-ball is path-connected (see point (1) in the appendix of [27]), as is the space of embeddings of a disk into a ball (see point (5)). The ambient space $\mathbb{R} \times I^2$ for a $\mathbb{Z}$-tangle can be decomposed along disks into a sequence of 3-balls, and $\mathbb{R} \times I^2$ itself is close enough to a 3-ball that the space of embeddings for each disk is also path-connected, so it follows that any boundary-fixing automorphism is isotopic to the identity.

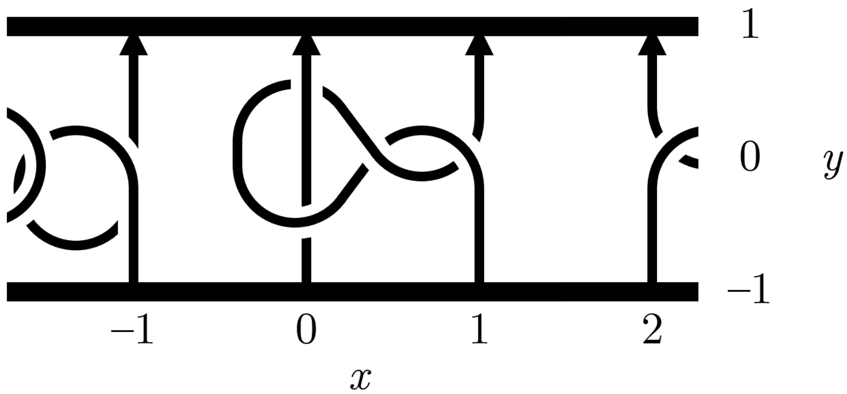

Figure 2.1: A generic $\mathbb{Z}$-tangle, shown only in part because it extends infinitely to the left and right. The z direction points out of the page toward the viewer.

There is also an action of $\mathbb{Z}$ on A on by translation: For $m \in \mathbb{Z}$ and $(x, y, z) \in \mathbb{R} \times I^2$, write $\tau^m(x, y, z) = (x + m, y, z)$. Then for $m \in \mathbb{Z}$ and $\ell \in A$, define $\tau^m \ell \in A$ (by abuse of notation) by

$$(\tau^m \ell)(n, t) \mapsto \tau^m(\ell(n - m, t)).$$

We can construct an action of $\mathbb{Z}$ on $\mathcal{I}$ similarly, and we will formally denote it with the redundant expression $(m, \iota) \mapsto \tau^m \iota \tau^{-m}$. Together, the three actions (of $\mathcal{I}$ on A, $\mathbb{Z}$ on A, and $\mathbb{Z}$ on $\mathcal{I}$) are related by the equation

$$(\tau^m \iota \tau^{-m}) \circ (\tau^m \ell) = \tau^m(\iota \circ \ell) \tag{2.1}$$

for all $m \in \mathbb{Z}$, $\iota \in \mathcal{I}$, and $\ell \in A$. We can also express this by saying we have an action of $\mathcal{I} \rtimes_\tau \mathbb{Z}$ on A. In the semidirect product, the visual tautology $\tau^m \iota \tau^{-m} = \tau^m \iota \tau^{-m}$ holds, so our notation is justified.

We will sometimes write just τ as shorthand for τ^1.

As a consequence of eq. (2.1), $\mathbb{Z}$ naturally acts on the orbit set $A/\mathcal{I}$ and the fixed set $\mathcal{I}_\mathbb{Z}$ naturally acts on the fixed set $A_\mathbb{Z}$.

We can express the relationships between A, $\mathcal{I}$, and $\mathbb{Z}$ in the following commutative diagram. Below, p and q are the natural quotient maps $A \twoheadrightarrow A/\mathcal{I}$ and $A_\mathbb{Z} \twoheadrightarrow A_\mathbb{Z}/\mathcal{I}_\mathbb{Z}$.

$$\begin{array}{ccccc}
A_{\mathbb{Z}} & \xhookrightarrow{\subset} & p^{-1}((A/\mathcal{I})_{\mathbb{Z}}) & \xhookrightarrow{\subset} & A \\
\big\downarrow q & & & & \\
A_{\mathbb{Z}}/\mathcal{I}_{\mathbb{Z}} & & \big\downarrow & & \big\downarrow p \\
\big\downarrow \mathrm{Bh} & & & & \\
p(A_{\mathbb{Z}}) & \xhookrightarrow{\subset} & (A/\mathcal{I})_{\mathbb{Z}} & \xhookrightarrow{\subset} & A/\mathcal{I}
\end{array} \tag{2.2}$$

Each element in the diagram can be described in words as follows:

- A is the set of $\mathbb{Z}$-tangles.
- $A/\mathcal{I}$ is the set of isotopy classes of $\mathbb{Z}$-tangles.
- $p^{-1}((A/\mathcal{I})_{\mathbb{Z}})$ is the set of $\mathbb{Z}$-tangles that are translation-invariant up to isotopy, i.e., those that are isotopic to all of their own translates.
- $(A/\mathcal{I})_{\mathbb{Z}}$ is the set of translation-invariant isotopy classes of $\mathbb{Z}$-tangles.
- $A_{\mathbb{Z}}$ is the set of translation-invariant $\mathbb{Z}$-tangles.
- $A_{\mathbb{Z}}/\mathcal{I}_{\mathbb{Z}}$ is the set of classes of translation-invariant $\mathbb{Z}$-tangles under translation-invariant isotopy. Later we will see how $A_{\mathbb{Z}}/\mathcal{I}_{\mathbb{Z}}$ is related to the set of knotoids.
- $p(A_{\mathbb{Z}})$ is the set of isotopy classes that have a translation-invariant representative.
- Bh sends each class of translation-invariant $\mathbb{Z}$-tangles up to translation-invariant isotopy to the (potentially) broader class up to general isotopy.

String Links

String links are a class of tangle-like objects defined in [26] for the purposes of studying link homotopy. We will give a definition here that shows their relationship to $\mathbb{Z}$-tangles.

Definition 2.1.2. For $k \in \mathbb{N}$, a *string link with k strings* is a ∂-proper embedding $\{1, \ldots, k\} \times I \to \mathbb{R} \times I^2$ with $(n, \pm 1) \mapsto (n, \pm 1, 0)$ for all $n \in \{1, \ldots, k\}$.[1]

The main distinguishing feature of a string link, in the context of tangles, is the correspondence between the "top" endpoints and "bottom" endpoints, reflected by the strings themselves.

[1]Some sources use a more general definition of "string link". In those sources, the objects herein would be called *pure string links*.

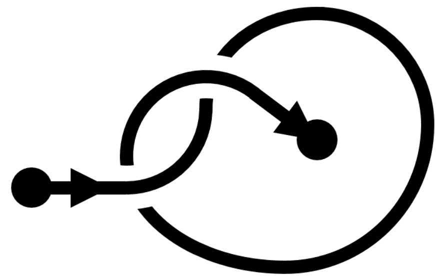

Figure 2.2: A knotoid diagram representing a “bifoil” knotoid. (As usual, we use the page as a proxy for S^2, considering there to be a “point at infinity”.)

Like $\mathbb{Z}$-tangles, string links are considered up to the action of $\mathcal{I}$. Unlike $\mathbb{Z}$-tangles, string links carry no natural notion of translation.

Of course, given a $\mathbb{Z}$-tangle ℓ and a $k \in \mathbb{N}$, we can create a string link with k strings by restricting ℓ to $\{1, \ldots, k\} \times I$ and forgetting the other strands. More generally, given ℓ and a subset of $\mathbb{Z}$ with k elements, there is a canonical-up-to-isotopy way to get a string link with k strands from the k strands in the subset: Everything can be pushed horizontally by varying amounts until each of the k strands has the correct endpoints.

2.2 Knotoids

There are several equivalent ways of defining knotoids, each with their own advantages. The traditional definitions are by knotoid diagrams and simple theta-curves, but later on we will mostly use the definition with annular 1-tangles. Here we will describe the different constructions and the relationships between them.

Diagrammatic Knotoids

Definition 2.2.1 (Turaev [49], Section 2.1)**.** A *knotoid diagram* consists of: (a) an immersion of $[0, 1]$ into S^2 with no self-intersections except for transverse double points away from the endpoints and (b) over/under crossing information at the self-intersections. A *knotoid* is an equivalence class of knotoid diagrams under ambient isotopy of S^2 and the Reidemeister moves performed away from the endpoints.

Note that knotoid diagrams are oriented. The images of 0 and 1 under the immersion of $[0, 1]$ are called the *tail* and the *head*, respectively.

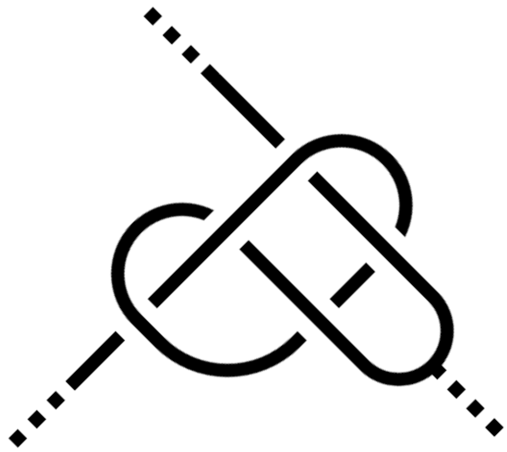

Figure 2.3: A slip knot. If the loop on the right is released, it becomes a mere unknot.

Simple Theta-Curves and Slip Knots

Definition 2.2.2 ([49], Section 5.1)**.** The *theta graph* Θ is the graph with two vertices, v_0 and v_1, and three oriented edges e_0, e_+, and e_- from v_0 to v_1. A *theta-curve* is an embedding of Θ into S^3. It is called *simple* if the image of $e_+ \cup e_-$ is an unknot.

Turaev defined a way of constructing simple theta-curves from knotoids: Given a knotoid diagram K, consider the ambient S^2 as being the "equator" of S^3 in the standard way. Our theta-curve will map v_0 and v_1 to the tail and head of K, respectively. Add e_+ and e_- on either side of S^2, each connecting v_0 to v_1 in an unknotted way, with e_+ "above" and e_- "below" S^2. Then consider the (possibly self-intersecting) path in S^2 defined by K. Let us embed e_0 into S^3 in such a way that it coincides with K everywhere except for in a neighborhood of each crossing, where it is "pushed off" in the appropriate direction.

Theorem 2.2.3 (Turaev [49], Section 5.1)**.** *The construction above provides a well-defined bijection between the set of knotoids and the set of isotopy classes of simple theta-curves.*

Knotoids and simple theta-curves can be viewed as mathematically capturing the real-life concept of a *slip knot*, which is essentially an unknot with a segment pulled away and held in place while the rest can move only in a restricted region. See, for example, Figure 2.3: If the rightmost dotted ray is ignored and the other two are used to extend the diagram to the point at infinity, then we have an unknot in S^3; however, if we consider only isotopies that leave the three dotted rays fixed, then we have a nontrivial object equivalent in structure to a simple theta-curve. (In the case shown, it corresponds to a bifoil knotoid.) This equivalence works because the ambient S^3 for a simple theta-curve can always be reparameterized relative to the position of v_0 and the initial directions of the three edges.

Annular 1-Tangles

Definition 2.2.4. An *annular* 1*-tangle* is a ∂-proper embedding of I into the solid torus $S^1 \times I^2$ that (a) maps ± 1 to $(0, \pm 1, 0)$ and (b) has winding number 0 around the S^1 factor.

The name "annular 1-tangle" is, of course, in reference to annular knots. Let $\overline{A}$ be the set of annular 1-tangles. Let $\overline{\mathcal{I}}$ be the set of boundary-fixing diffeomorphisms of $S^1 \times I^2$.

Of course, we can canonically identify $\overline{A}$ with $A_{\mathbb{Z}}$ and $\overline{\mathcal{I}}$ with $\mathcal{I}_{\mathbb{Z}}$. Specifically, for $\overline{\ell} \in \overline{A}$ and $n \in \mathbb{Z}$, there is a unique lift of $\overline{\ell}$ to a map $I \to \mathbb{R} \times I^2$ that maps ± 1 to $(n, \pm 1)$, so we let $s(\overline{\ell}) \in A_{\mathbb{Z}}$ be the translation-invariant $\mathbb{Z}$-tangle that restricts to $\{n\} \times I$ as the n^{th} lift. Similarly, each element of $\overline{\mathcal{I}}$ lifts to an automorphism of $\mathbb{R} \times I^2$, and there is a unique lift that restricts to the identity on the boundary. We can extend the diagram in Equation (2.2) to the commutative diagram below.

$$
\begin{array}{ccccccc}
\overline{A} & \overset{s}{\hookrightarrow\!\!\!\rightarrow} & A_{\mathbb{Z}} & \overset{\subset}{\hookrightarrow} & p^{-1}((A/\mathcal{I})_{\mathbb{Z}}) & \overset{\subset}{\hookrightarrow} & A \\
\downarrow\!\!\downarrow \scriptstyle{\overline{q}} & & \downarrow\!\!\downarrow \scriptstyle{q} & & \downarrow\!\!\downarrow & & \downarrow\!\!\downarrow \scriptstyle{p} \\
\overline{A}/\overline{\mathcal{I}} & \overset{s_*}{\hookrightarrow\!\!\!\rightarrow} & A_{\mathbb{Z}}/\mathcal{I}_{\mathbb{Z}} & & & & \\
 & & \downarrow\!\!\downarrow \scriptstyle{\mathrm{Bh}} & & & & \\
 & & p(A_{\mathbb{Z}}) & \overset{\subset}{\hookrightarrow} & (A/\mathcal{I})_{\mathbb{Z}} & \overset{\subset}{\hookrightarrow} & A/\mathcal{I}
\end{array}
\tag{2.3}
$$

A correspondence between knotoids and objects similar to annular 1-tangles is noted in, for example, [48] (page 2). We will make that explicit here by showing that there is a natural way of equating $\overline{A}/\overline{\mathcal{I}}$ with classes of simple theta-curves:

$$\{\text{knotoids}\} \overset{[49]}{\cong} \{\text{simple theta-curves}\}/\text{isotopy} \overset{?}{\cong} \overline{A}/\overline{\mathcal{I}} \cong A_{\mathbb{Z}}/\mathcal{I}_{\mathbb{Z}}$$

We can identify S^3 with the quotient of $S^1 \times I^2$ by the relation $(x_1, y, z) \sim (x_2, y, z)$ for all $x_1, x_2 \in S^1$ and $(y, z) \in \partial I^2$. Given some $\overline{\ell} \in \overline{A}$, let us construct a simple theta-curve $\theta_{\overline{\ell}}$ as follows: The composition of $\overline{\ell}$ with the quotient map $S^1 \times I^2 \to S^3$ above defines the embedding of e_0. Then we define e_+ (resp. e_-) to coincide with the image of $\{(x, y, z) \in S^1 \times \partial I^2 \mid z > 0\}$ (resp. $z < 0$) under the quotient map. By [27] (point (9) in the appendix), the isotopy class of the resulting theta-curve is invariant under the action of $\overline{\mathcal{I}}$ on $\overline{\ell}$, so we have a well-defined map $\overline{A}/\overline{\mathcal{I}} \to \{\text{simple theta-curves}\}/\text{isotopy}$.

Proposition 2.2.5. *The map $\overline{A}/\overline{\mathcal{I}} \to \{$simple theta-curves$\}/$isotopy above is a bijection.*

Proof. For surjectivity, suppose that we are given a simple theta-curve θ. Since $\theta(e_+ \cup e_-)$ is an unknot, and since we only care about θ up to isotopy, we can assume that θ behaves appropriately on e_+ and e_- for being of the form $\theta_{\overline{\ell}}$. Now let us consider the preimage of $\theta(e_0)$ under the quotient $S^1 \times I^2 \to S^3$. We want the strand to have the correct endpoints and winding number to make an element of $\overline{A}$. We can achieve this by isotoping θ in a neighborhood of v_0 and v_1, twirling each of them to point in the right directions and then possibly more to get the correct winding number. Once we have done this, θ is of the form $\theta_{\overline{\ell}}$.

Now, for injectivity, suppose we have $\overline{\ell}$ and $\overline{\ell}'$ such that $\theta_{\overline{\ell}}$ is isotopic to $\theta_{\overline{\ell}'}$. Then there is a diffeomorphism $\varphi\colon S^3 \to S^3$ relating $\theta_{\overline{\ell}}$ to $\theta_{\overline{\ell}'}$. The restriction of φ to $\theta_{\overline{\ell}}(e_+ \cup e_-)$ is the identity. Consider the derivatives of φ at each point on this unknot. The derivatives in the normal direction define a framing for both arcs, and the stipulation that $\overline{\ell}$ and $\overline{\ell}'$ each have zero winding number implies that the framing is the same as it would be if the derivatives were all the identity maps on their respective tangent spaces. Therefore we can perturb φ in a neighborhood of that circle, without affecting the property that φ sends $\theta_{\overline{\ell}}$ to $\theta_{\overline{\ell}'}$, to make it so that the derivatives at each point on the circle are all the identity maps. Therefore, φ lifts to a boundary-preserving diffeomorphism of $S^1 \times I^2$. □

The relationship between an isotopy class of annular 1-tangles and the corresponding knotoid can be understood fairly easily without involving simple theta-curves: Given a $\overline{\ell} \in \overline{A}$, we can form a diagram of $\overline{\ell}$ by composing $\overline{\ell}$ with the natural projection $S^1 \times I^2 \to S^1 \times I$ that forgets the z (third) coordinate to obtain a path on an annulus. The path will start at $(0, -1)$ and end at $(0, 1)$. When $\overline{\ell}$ is in general position, the path on the annulus will be immersed and only self-intersect transversely with no triple points. We can then create a knotoid diagram by adding crossing information (where greater z values are "over" lesser z values) and placing $S^1 \times I$ onto S^2 in an orientation-preserving way that sends each boundary component to a single point and is otherwise a diffeomorphism.

2.3 $\mathbb{Z}$-Braids and $\mathbb{N}$-Braids

Definition 2.3.1. A $\mathbb{Z}$-*braid* is an automorphism of $\mathbb{R} \times I$ that (a) fixes the boundary pointwise and (b) preserves $\mathbb{Z} \times \{0\}$ as a set. A $\mathbb{Z}$-braid is *pure* if it fixes $\mathbb{Z} \times \{0\}$ pointwise.

Let B denote the topological group of pure $\mathbb{Z}$-braids. Let us write $b_1 \sim b_2$ whenever b_1 and b_2 are in the same component, that is, if b_1 and b_2 are isotopic through elements of B. Thus the quotient group $B/{\sim}$ is essentially a pure mapping class group. It differs from standard mapping class groups, even big mapping class groups, in that the surface in question doesn't have compact boundary. (See [17].) We we denote the quotient function $B \mapsto B/{\sim}$ simply with $b \mapsto [b]$.

As in the case of $\mathbb{Z}$-tangles, there is an action of $\mathbb{Z}$ on B by translation, which we will also denote by τ. The translation action descends to an action on $B/{\sim}$. On the level of isotopy classes, applying τ is the same as a conjugating by a certain non-pure $\mathbb{Z}$-braid.

When dealing with $\mathbb{Z}$-braids, we will use the coordinates x and z, in that order, for points in $\mathbb{R} \times I$.

Later we will also allude to a similar concept on an annulus: An *annular* 1-*braid* is an automorphism of $S^1 \times I$ that (a) fixes $(S^1 \times \{-1, 1\}) \cup \{(0, 0)\}$ pointwise and (b) induces the trivial automorphism on the relative first homology group $H_1(S^1 \times I, \partial \cup \{(0,0)\})$. Let $\overline{B}$ be the group of annular 1-braids. We can identify $\overline{B}$ with the fixed set $B_{\mathbb{Z}}$. By capping each boundary component, we have a map from $\overline{B}$ to the set of pure (orientation-preserving) automorphisms of a thrice-punctured sphere. On the level of isotopy classes, this map is a group isomorphism. Therefore, $\overline{B}$ actually has only one component, so the annular 1-braid group is trivial.

If we replace $\mathbb{Z}$ in Definition 2.3.1 with $\mathbb{N}$, we obtain objects equivalent to those studied in [18]:

Definition 2.3.2. A $\mathbb{N}$-*braid* is an automorphism of $\mathbb{R} \times I$ that (a) fixes the boundary pointwise and (b) preserves $\mathbb{N} \times \{0\}$ as a set. A $\mathbb{N}$-braid is *pure* if it fixes $\mathbb{N} \times \{0\}$ pointwise.

Isotopy between $\mathbb{N}$-braids is allowed to move points outside of $(\mathbb{R} \times \{\pm 1\}) \cup (\mathbb{N} \times \{0\})$. The set of pure $\mathbb{N}$-braids up to isotopy naturally carries an operation τ^{-1} that forgets the first strand and shifts all of the others left by 1:

$$(\tau^{-1}b)(x,z) = \tau^{-1}(b(x+1,z))$$

Despite the notation, this τ^{-1} is non-invertible. Just like the τ for $\mathbb{Z}$-tangles, τ^{-1} descends to isotopy classes.

The Alexander Method

For each $n \in \mathbb{Z}$, let $\gamma_n = \{n\} \times [-1,0]$ and $\delta_n = \{n\} \times [0,1]$. Using the γ_n and δ_n, we will see that an "Alexander method" holds for $\mathbb{Z}$-braids. For each $b \in B$, the image $b(\gamma_n)$ of each γ_n under b is an embedded path with the same endpoints as γ_n.

Proposition 2.3.3. *Suppose $b \in B$ has the property that, for all n, $b(\gamma_n)$ is homotopic to γ_n through an endpoint-fixing family of paths that otherwise avoid $\mathbb{Z} \times \{0\}$. Then b is isotopic to the identity.*

Proof. Our proof is similar to the proof of Lemma 2.5 in [19], but we have to be careful about the lack of compactness. For every individual n, the homotopy from $b(\gamma_n)$ to γ_n can be promoted to an isotopy ([19], cf. Proposition 1.10, Section 1.2.7).

We will define a sequence $(b_i), i \in \{0\} \cup \mathbb{N}$, of elements of B as well as chosen isotopies between them. In particular, each b_i will leave γ_n fixed for $n \in \mathbb{Z}$ with $|n| \leq i$. Furthermore, any compact subset of $\mathbb{R} \times I$ will, for large i, remain fixed by the chosen isotopy from b_{i-1} to b_i throughout the isotopy.

Let us start by isotoping b to a new $b_0 \in B$ such that b_0 fixes γ_0 pointwise. Now, as a recursive step, for any $i \in \mathbb{N}$, once we have fixed b_{i-1}, let us define a new $b_i \in B$ and our isotopy from b_{i-1} to b_i: Consider that $b_{i-1}(\gamma_i)$ is homotopic to γ_i and disjoint from all γ_n for $|n| \leq i$. Since removing those γ_n is algebraically equivalent to deleting a 1-cell, $b_{i-1}(\gamma_i)$ is homotopic to γ_i in the surface formed by deleting those γ_n. Then it is also isotopic in the surface with those paths cut out. We can make the same argument about cutting along vertical lines slightly to either side of the left- and right-most points of $b_{i-1}(\gamma_i)$. Therefore, let us choose an isotopy from b_{i-1} to a new pure $\mathbb{Z}$-braid that fixes γ_i in such a way that the isotopy doesn't affect anything more than slightly to the left or right of $b_{i-1}(\gamma_i)$. Following this up with another isotopy to deal with γ_{-1}, we obtain our b_i and chosen isotopy from b_{i-1} to b_i.

The paths γ_n together form an end-properly embedded subset of $\mathbb{R} \times I$, and any compact region in $\mathbb{R} \times I$ only contains parts of finitely many of the γ_n. Any automorphism b must preserve this property. Therefore, our chosen sequence of isotopies has the desired property that any given compact region will eventually stay fixed. Hence we are allowed to take the infinite concatenation of the isotopies to get an isotopy from b to a new $b_\infty \in B$ that fixes γ_n for all $n \in \mathbb{Z}$.

Now we will find an isotopy from b_∞ to the identity using a similar strategy as above but applied to the δ_n. The complement of all of the γ_n is an infinite disk. Holding all of the γ_n fixed, let us isotope b_∞ to a new braid that also fixes δ_0. Then we can move $\delta_{\pm 1}$ into position, followed by everything in between using the usual Alexander lemma ([19], Lemma 2.1). Continuing with $\delta_{\pm 2}$, and so on, we get the desired isotopy. □

2.4 Spanning Disks and Height

Here we will define objects called "spanning disks" in the separate contexts of both $\mathbb{Z}$-tangles and annular 1-tangles.

$\mathbb{Z}$-Tangles

Definition 2.4.1. A *spanning disk* for a $\mathbb{Z}$-tangle $\ell \in A$ is a ∂-properly embedded disk $D \subset \mathbb{R} \times I^2$ with the following properties:

- The composite map
$$\partial D \subset \partial(\mathbb{R} \times I^2) = \mathbb{R} \times \partial I^2 \twoheadrightarrow \partial I^2,$$
where the last arrow is the projection onto the second factor, is a diffeomorphism.[2]
- The intersections between D and ℓ are transverse and occur away from the boundary.

Given a $\mathbb{Z}$-tangle ℓ and a spanning disk D, we can cut $\mathbb{R} \times I^2$ along D to get a space with two components. Let $R_+(D)$ be the component of $(\mathbb{R} \times I^2) \setminus D$ in which the x values are bounded below, and let $R_-(D)$ be the one in which the x values are bounded above. The *height* of the pair (ℓ, D), written $H(\ell, D)$, is defined to be the number of intersections between ℓ and D. Each such intersection is signed, because D has a canonical orientation: An intersection of ℓ with D is *positive* if ℓ is oriented into $R_+(D)$, and it is *negative* otherwise. (See Figure 2.4) The *positive* (resp. *negative*) *height* of (ℓ, D), denoted $H_+(\ell, D)$ (resp. H_-), is the number of positive (resp. negative) intersections.

[2] Here, of course, we aren't using "diffeomorphism" in the strictest possible sense. Recall Section 1.3.

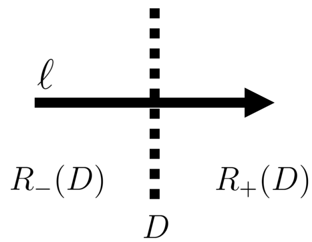

Figure 2.4: A positive intersection between ℓ and a spanning disk D.

Given just ℓ, the *height* $h(\ell)$ is the minimum of $H(\ell, D)$ over all choices of D. Of course, h is invariant under the action of $\mathcal{I}$. We define the *positive* and *negative heights* similarly. Usually we will only care about the heights of $\mathbb{Z}$-tangles ℓ satisfying $p(\ell) \in (A/\mathcal{I})_{\mathbb{Z}}$.

Note that any spanning disk D can have its boundary pushed (without affecting anything near ℓ or outside a neighborhood of $\partial(\mathbb{R} \times I^2)$ in such a way that, after the movement, ∂D will constitute the intersection between $\partial(\mathbb{R} \times I^2)$ and some plane in $\mathbb{R}^3$ parallel to the z axis. To be even more rigid, we could make it so that the x values of the points of ∂D on $\mathbb{R} \times \{\pm 1\} \times I$ are half-integers.

Annular 1-Tangles

The concepts of spanning disks and height transfer from $\mathbb{Z}$-tangles to annular 1-tangles with few modifications.

Definition 2.4.2. A *spanning disk* for an annular 1-tangle $\overline{\ell} \in \overline{A}$ is a ∂-properly embedded disk $\overline{D} \subset S^1 \times I^2$ with the following properties:

- The composite map
$$\partial\overline{D} \subset \partial(S^1 \times I^2) = S^1 \times \partial I^2 \twoheadrightarrow \partial I^2,$$
where the last arrow is the projection onto the second factor, is a diffeomorphism.
- The intersections between $\overline{D}$ and $\overline{\ell}$ are transverse and occur away from the boundary.

Given an annular 1-tangle $\overline{\ell}$ and a spanning disk $\overline{D}$, the *height* of the pair $(\overline{\ell}, \overline{D})$, written $\overline{H}(\overline{\ell}, \overline{D})$, is defined to be the number of intersections between $\overline{\ell}$ and $\overline{D}$. Each such intersection is signed, because $\overline{D}$ has a canonical orientation: For $\overline{D}$ of the form $\{\text{pt}\} \times I^2$, an intersection of $\overline{\ell}$ with $\overline{D}$ is *positive* if $\overline{\ell}$ is oriented in such a way that higher t values correspond with higher x values, and it is *negative* otherwise; for spanning disks of other forms we extend the previous definition continuously. The *positive* (resp. *negative*) *height* of $(\overline{\ell}, \overline{D})$, denoted $\overline{H}_+(\overline{\ell}, \overline{D})$ (resp. $\overline{H}_-$), is the number of positive (resp. negative) intersections.

Since the intersections between $\overline{\ell}$ and $\overline{D}$ each correspond to a point on I, we can also consider the order of the types of intersections. A *sign sequence* is a finite sequence with values in $\{+,-\}$. We define $\mathrm{Seq}(\bar{\ell},\overline{D})$ to be the sign sequence with $\overline{H}(\bar{\ell},\overline{D})$ terms expressing the signs of the intersections in order from -1 to 1.

Given just $\bar{\ell}$, the *height* $\overline{h}(\bar{\ell})$ is the minimum of $\overline{H}(\bar{\ell},\overline{D})$ over all choices of $\overline{D}$. Of course, $\overline{h}$ is invariant under the action of $\overline{\mathcal{I}}$. We define the *positive* and *negative heights* similarly. A sign sequence will be called *attainable for* $\bar{\ell}$ if it is of the form $\mathrm{Seq}(\bar{\ell},\overline{D})$ for some spanning disk $\overline{D}$.

As in the case of $\mathbb{Z}$-tangles, for the purposes of height and attainable sign sequences a spanning disk $\overline{D}$ can always be taken to have $\overline{D}\cap(S^1\times\{\pm1\}\times I)=\{\frac{1}{2}\}\times\{\pm1\}\times I$.

Knotoids

The original definition of height appeared in [49] (Section 4.3) under the name *complexity* and was applied to diagrammatic knotoids. The height of a knotoid can be defined in a manner very similar to the height of an annular 1-tangle, using *shortcuts* (see below) in place of spanning disks.

Definition 2.4.3 ([49], Section 2.2)**.** A *shortcut* for a knotoid diagram K is an embedded path in S^2 from the tail to the head that intersects K transversely and does not intersect the crossings.

We can define the height of a knotoid diagram K to be the minimum of the heights of (K,a) over all shortcuts a for K. However, the height of a knotoid diagram is not invariant under the Reidemeister moves. Therefore the height of a knotoid (that is, not just of a knotoid diagram) is defined as the minimum of the heights of of all representative diagrams. For this reason, annular 1-tangles and simple theta-curves are easier to work with than knotoids for the purposes of talking about height.

Although [49] only addresses unsigned heights, intersections between a knotoid diagram and a shortcut can still be considered to be signed, leading to concepts of signed heights and attainable sign sequences for knotoids, depending on the exact convention used to specify signs of shortcut intersections.

It is a straightforward exercise to convince oneself that the height of a knotoid coincides with the height of the corresponding class of annular 1-tangles under the equivalence described in Section 2.2. The same goes for attainable sign sequences.

2.5 Other Important Functions

The Birman–Hilden Map

The forgetful map Bh in eq. (2.2) can be viewed as an invariant of annular 1-tangles when composed with the bijection s_* in eq. (2.3). See Figure 2.6. We will call it the *Birman–Hilden map* in reference to the Birman–Hilden theory of mapping class groups (see [11, 9, 10], also [39] for an exposition).

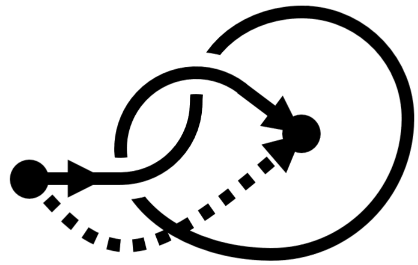

Figure 2.5: A shortcut for a bifoil knotoid diagram. The sign sequence of this diagram and shortcut is $(-)$.

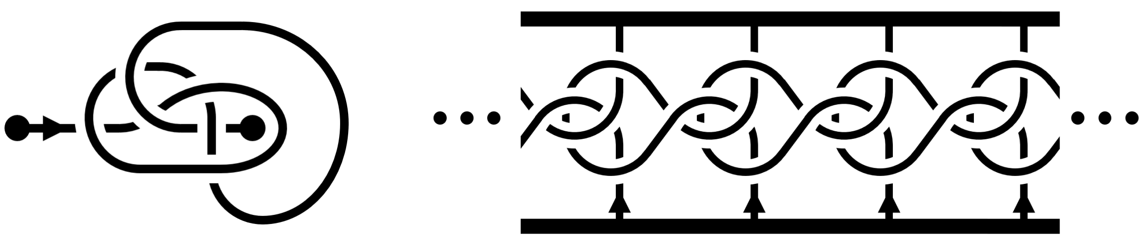

Figure 2.6: A knotoid diagram (left) and corresponding translation-invariant $\mathbb{Z}$-tangle (right).

We will address the Birman–Hilden invariant more in later chapters.

Basic Involutions

Several operations defined for knotoids in [49] (Section 3.2) easily transfer to annular 1-tangles and $\mathbb{Z}$-tangles:

For a $\mathbb{Z}$-tangle $\ell \in A$ or annular 1-tangle $\overline{\ell} \in \overline{A}$, the *mirror image* $\mathrm{mir}(\ell) \in A$ or $\mathrm{mir}(\overline{\ell}) \in \overline{A}$ is obtained by flipping the z direction. That is, for example, if $\ell(n,t) = (x,y,z)$ for some n, t, x, y, and z, then $\mathrm{mir}(\ell)(n,t) = (x,y,-z)$. The *symmetry* operation, denoted $\ell \mapsto \mathrm{sym}(\ell)$ and $\overline{\ell} \mapsto \mathrm{sym}(\overline{\ell})$, flips in the x direction. Specifically, for $\mathbb{Z}$-tangles, if ℓ is as above then $\mathrm{sym}(\ell)(-n,t) = (-x,y,z)$. *Rotation* is the composition of symmetry and mirror image reflection. Lastly, the *reverse* of a $\mathbb{Z}$-tangle or annular 1-tangle is obtained by flipping in the x and y directions, using the equation $\mathrm{rot}(\ell)(-n,-t) = (-x,-y,z)$.

It is clear that these operations, together with their various compositions, constitute eight mutually-commuting involutions on A and $\overline{A}$ (including one trivial one). Height, but not signed height, is invariant under these involutions.

The two versions of each involution (for $\mathbb{Z}$-tangles and annular 1-tangles) are related by composition with s. They also descend to $A/\mathcal{I}$ and $\overline{A}/\overline{\mathcal{I}}$ and factor through Bh.

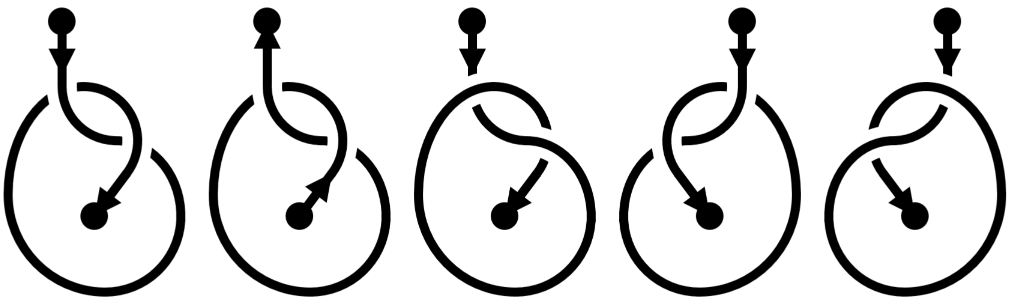

Figure 2.7: From left to right, a diagram for "bifoil" knotoid k, $\text{rev}(k)$, $\text{mir}(k)$, $\text{sym}(k)$, and $\text{rot}(k)$. Only the first two are equal as knotoids.

Diffeomorphisms

Let $\mathcal{J}$ be the set of orientation-preserving diffeomorphisms from $\mathbb{R} \times I^2$ to itself that fix $\mathbb{Z} \times \{\pm 1\} \times \{0\}$ pointwise. Then $\mathcal{I}$ is a normal subgroup of $\mathcal{J}$ and the action of $\mathcal{I}$ on A extends to an action of $\mathcal{J}$ in a natural way. Furthermore, the translation action of $\mathbb{Z}$ on $\mathcal{I}$ extends to an action on $\mathcal{J}$. All together, we have an action of $\mathcal{J} \rtimes_\tau \mathbb{Z}$ on A. This gives us actions of $\mathcal{J}/\mathcal{I}$ on $A/\mathcal{I}$ and of $\mathcal{J}_\mathbb{Z}$ on $A_\mathbb{Z}$.

Similarly, let $\overline{\mathcal{J}}$ be the set of orientation-preserving diffeomorphisms of $S^1 \times I^2$ that (a) fix $\{0\} \times \{\pm 1\} \times \{0\}$ pointwise and (b) induce the trivial automorphism on the relative first homology group $H_1(S^1 \times I^2, \{0\} \times \{\pm 1\} \times \{0\})$. Then $\overline{\mathcal{I}}$ is a normal subgroup of $\overline{\mathcal{J}}$ and the action of $\overline{\mathcal{I}}$ on $\overline{A}$ can be extended to $\overline{\mathcal{J}}$. As a result we have an action of $\overline{\mathcal{J}}/\overline{\mathcal{I}}$ on $\overline{A}/\overline{\mathcal{I}}$.

We can canonically identify $\overline{\mathcal{J}}$ with $\mathcal{J}_\mathbb{Z}$, and under this identification the action of $\overline{\mathcal{J}}$ on $\overline{A}$ matches that of $\mathcal{J}_\mathbb{Z}$ on $A_\mathbb{Z}$.

The action of $\mathcal{J}/\mathcal{I}$ and $\overline{\mathcal{J}}/\overline{\mathcal{I}}$ on $A/\mathcal{I}$ and $\overline{A}/\overline{\mathcal{I}}$ can be though of as extending the basic involutions above. In particular, the basic involutions can be viewed as coming from particular automorphisms of $\mathbb{R} \times I^2$ or $S^1 \times I^2$ that don't follow all of the conditions for being in $\mathcal{J}$ or $\overline{\mathcal{J}}$. In the case of $\overline{A}/\overline{\mathcal{I}}$, the action of $\overline{\mathcal{J}}/\overline{\mathcal{I}}$ together with the basic involutions can all be considered to be special cases of a single action by $\overline{\mathcal{J}} \rtimes (\mathbb{Z}/2\mathbb{Z})^3$. Unlike the basic involutions, the actions of $\mathcal{J}$ and $\overline{\mathcal{J}}$ don't preserve height.

Spanning Disk Framings

For a given annular 1-tangle $\overline{\ell} \in \overline{A}$ and a spanning disk $\overline{D}$, the algebraic intersection number between $\overline{\ell}$ and $\overline{D}$ is an integer that only depends on $\overline{D}$. An equivalence class of spanning disks for $\overline{\ell}$ with the same algebraic intersection number can be called a *spanning disk framing* for $\overline{\ell}$. The counterpart concept for knotoids also appears in [42] (Section 2.2.3) under the name *coframed knotoids* and is related to the *Morse knotoids* of [25] (Definition 2).

Of course, a disk-framed annular 1-tangle could simply be thought of as an annular 1-tangle paired with an arbitrary integer. However, another way to recover the same concept would be to alter the definition of annular 1-tangles to remove the winding number condition. Doing so would have multiple benefits: Besides slightly simplifying the definition of annular 1-tangles, it would allow us to define heights and sign sequences using only spanning disks of the form $\{1/2\} \times I^2$. Furthermore, the action of $\overline{\mathcal{J}} \rtimes (\mathbb{Z}/2\mathbb{Z})^3$ naturally carries over to these disk-framed annular 1-tangles, and this extension is nontrivial in that a single element of $\overline{\mathcal{J}} \rtimes (\mathbb{Z}/2\mathbb{Z})^3$ applied to two different framings of the same annular 1-tangle can give back two framed annular 1-tangles that are distinct even after forgetting their framings.

Oriented Knots

An important subclass of annular 1-tangles and of $\mathbb{Z}$-tangles, originally defined for knotoids in [49] (Section 3.3), are those that are *knot-type*. For every oriented knot isotopy class κ we can easily construct a well-defined element $\kappa^\bullet$ of $\overline{A}/\overline{\mathcal{I}}$ as follows: Start with the annular 1-tangle defined by $\overline{\ell}(t) = (0, t, 0)$, then tie κ into a small neighborhood of the strand. An element of $\overline{A}/\overline{\mathcal{I}}$ is *knot-type* if it can be constructed as $\kappa^\bullet$ in this way. We will similarly call a specific annular 1-tangle knot-type if it represents a knot-type class. An annular 1-tangle that is not knot-type is *proper*. Of course, each knot-type annular 1-tangle has height 0, and the converse is true as well: If $\overline{\ell}$ has height 0, then consider a spanning disk $\overline{D}$ realizing the height. Cutting $S^1 \times I^2$ along $\overline{D}$ results in a long knot interpretable as an oriented knot. The class of $\overline{\ell}$ is obtainable from this same oriented knot, because an isotopy can put $\overline{D}$ into the position of $\{1/2\} \times I^2$.

On the other side, a $\mathbb{Z}$-tangle $\ell \in A$ is *knot-type* is there is a knot-type $\overline{\ell} \in \overline{A}$ with $p(\ell) = p(s(\overline{\ell}))$. We can also apply the term knot-type to $p(\ell)$. Every knot-type $\mathbb{Z}$-tangle has height 0, but the converse does not hold. We will see in Section 6.1 that even some elements of $(A/\mathcal{I})_{\mathbb{Z}}$ with height 0 are not knot-type.

Two distinct isotopy classes of oriented knots have distinct corresponding knot-type isotopy classes of annular 1-tangles. One easy way to see this is by the *over-* and *underpass closures* defined in terms of diagrammatic knotoids in [49] (Section 3.3). From a simple theta-curve θ, we can obtain two oriented knots from $\theta(e_0 \cup e_+)$ and $\theta(e_0 \cup e_-)$ by using the orientations provided by e_0. Thus from an annular 1-tangle $\overline{\ell}$, the over- and underpass closures, respectively, are the well-defined oriented knots obtained in those ways from $\theta_{\overline{\ell}}$. For a knot-type $\overline{q}(\overline{\ell}) = \kappa^\bullet \in \overline{A}$, the over- and under-pass closures of $\overline{\ell}$ recover κ.

Multiplication

Given $\mathbb{Z}$-tangles ℓ_1 and ℓ_2, we define their product $\ell_1\ell_2$ by stacking them, with ℓ_1 compressed into $\mathbb{R} \times [-1, 0] \times I$ and ℓ_2 compressed into $\mathbb{R} \times [0, 1] \times I$. In symbols, we have the following three formulae for the x, y, and z coordinates of $(\ell_1\ell_2)(n,t)$:

$$(\ell_1\ell_2)(n,t)_x = \begin{cases} \ell_1(n,2t+1)_x & -1 \le t \le 0 \\ \ell_2(n,2t-1)_x & 0 \le t \le 1 \end{cases}$$

$$(\ell_1\ell_2)(n,t)_y = \begin{cases} (\ell_1(n,2t+1)_y - 1)/2 & -1 \le t \le 0 \\ (\ell_2(n,2t-1)_y + 1)/2 & 0 \le t \le 1 \end{cases}$$

$$(\ell_1\ell_2)(n,t)_z = \begin{cases} \ell_1(n,2t+1)_z & -1 \le t \le 0 \\ \ell_2(n,2t-1)_z & 0 \le t \le 1 \end{cases}$$

Using the same set of formulae, we can also define multiplication the set of string links with k strings in a way that coincides with the definitions in [26] (page 397) and [12] (page 3), and on $\overline{A}$ in a way that coincides with the definition in [49] (Section 4.1). It is straightforward to check that multiplication is well-defined for $A_{\mathbb{Z}}$, $A/\mathcal{I}$, $\overline{A}/\overline{\mathcal{I}}$, etc. The multiplication operations on $\overline{A}$ and $A_{\mathbb{Z}}$ are compatible with s, and for two oriented knots κ_1 and κ_2 we have $(\kappa_1\kappa_2)^\bullet = \kappa_1^\bullet\kappa_2^\bullet$.

As binary operations on $\overline{A}/\overline{\mathcal{I}}$ and $A/\mathcal{I}$, multiplication is a monoid, with the knot-type annular 1-tangle and $\mathbb{Z}$-tangle corresponding to the unknot serving as respective identities. In $A/\mathcal{I}$, the action of $\mathbb{Z}$ is a monoid homomorphism.

A *prime* annular 1-tangle is one that cannot be written as a nontrivial product. A number of fundamental results about multiplication of knotoids show up in [49] (Section 4.2); their equivalents for annular 1-tangles are as follows: Every annular 1-tangle has a unique decomposition of the form

$$\kappa^\bullet\overline{\ell}_1\overline{\ell}_2\cdots\overline{\ell}_n,$$

where each $\overline{\ell}_i$ is a proper prime annular 1-tangle. A knot-type annular 1-tangle is prime if and only if the corresponding knot is prime. Two distinct prime annular 1-tangles commute if and only if one or both is knot-type.

Definition 2.5.1. A *decomposing annulus* for an annular 1-tangle $\overline{\ell}$ is a ∂-properly embedded surface $\overline{Q} \subset S^1 \times I^2$ with the following properties:

- $\overline{Q}$ is diffeomorphic to $S^1 \times I$.
- $\partial\overline{Q} = S^1 \times \{0\} \times \{-1,1\}$.
- $\overline{Q}$ is transverse to ℓ.
- $\overline{Q}$ intersects ℓ only once.

An annular 1-tangle $\overline{\ell}$ can be written as a product (on the nose, not just up to isotopy) if and only if $S^1 \times \{0\} \times I$ is a decomposing annulus for $\overline{\ell}$. More generally, if $\overline{Q}$ is a decomposing annulus for an annular 1-tangle $\overline{\ell}$, then there are annular 1-tangles $\overline{\ell}_1, \overline{\ell}_2 \in \overline{A}$ and $\overline{\iota} \in \overline{I}$ such that $\overline{\ell} = \overline{\iota} \circ (\overline{\ell}_1\overline{\ell}_2)$ and $\overline{Q} = \overline{\iota}(S^1 \times \{0\} \times I)$. The factor annular 1-tangles $\overline{\ell}_1, \overline{\ell}_2$ are uniquely determined up to isotopy, because the only choice in choosing the decomposition amounts to picking an element of $\overline{B}$, but as we saw in Section 2.3, the annular 1-braid group is trivial.

The situation is slightly different in the context of $\mathbb{Z}$-tangles:

Definition 2.5.2. A *decomposing ribbon* for a $\mathbb{Z}$-tangle ℓ is a biproperly embedded surface $Q \subset \mathbb{R} \times I^2$ with the following properties:

- Q is diffeomorphic to $\mathbb{R} \times I$.
- $\partial Q = \mathbb{R} \times \{0\} \times \{-1, 1\}$.
- Q is transverse to ℓ.
- Q intersects each strand of ℓ only once.

As with annular 1-tangles, a $\mathbb{Z}$-tangle ℓ can be written as a product if and only if $\mathbb{R}\times\{0\}\times I$ is a decomposing ribbon for ℓ. Also, if Q is a decomposing ribbon for a $\mathbb{Z}$-tangle ℓ, then there are $\mathbb{Z}$-tangles $\ell_1, \ell_2 \in A$ and $\iota \in \mathcal{I}$ such that $\ell = \iota \circ (\ell_1 \ell_2)$ and $Q = \iota(\mathbb{R} \times \{0\} \times I)$. However, a decomposing ribbon does not uniquely determine the factor $\mathbb{Z}$-tangles. In Section 4.3 we will show that we can make such a statement in the restricted context of $(A/\mathcal{I})_{\mathbb{Z}}$.

Lifting

Definition 2.5.3. For a $\mathbb{Z}$-tangle $\ell \in A$ and some $k \in \mathbb{N}$, we define the k^{th} *lift* of ℓ to be the $\mathbb{Z}$-tangle ℓ/k formed by ignoring all strands of ℓ not corresponding to a multiple of k and then compressing the remaining ones in the x direction by a factor of k. In formulae, the components of ℓ/k are as follows:

$$(\ell/k)(n,t)_x = \ell(kn,t)_x/k$$
$$(\ell/k)(n,t)_y = \ell(kn,t)_y$$
$$(\ell/k)(n,t)_z = \ell(kn,t)_z$$

Lifting is, of course, compatible with multiplication of $\mathbb{Z}$-tangles and induces operations on isotopy classes as well. Therefore we will be comfortable writing, for example, $p(\ell)/k \in A_{\mathcal{I}}$. The lift of a translation-invariant $\mathbb{Z}$-tangle is also translation-invariant.

Lifting satisfies a formula $\ell/(k_1 k_2) = (\ell/k_1)/k_2$ for $\ell \in A$ and $k_1, k_2 \in \mathbb{N}$. For a fixed $\ell \in A_{\mathbb{Z}}$, the sequence $(p(\ell/k))$ stabilizes for large k in a knot-type isotopy class of $\mathbb{Z}$-tangles corresponding to the oriented knot formed just by considering what happens with a single strand of ℓ. We will call this knot $p(\ell)/\infty$.

The lifting operations on $A_{\mathbb{Z}}$ induce corresponding operations on $\overline{A}$ (and, by extension, $\overline{A}/\overline{\mathcal{I}}$). The formulae used in Definition 2.5.3 do not carry over, since division by natural numbers is not defined in S^1. Instead, the best way to think about a the k^{th} lift of an annular 1-tangle $\overline{\ell}$ is as the result of choosing a lift (in the standard topological sense) of $\overline{\ell}$

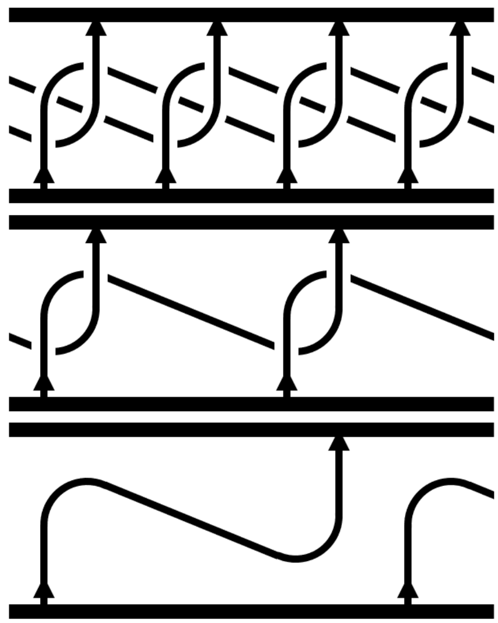

Figure 2.8: A $\mathbb{Z}$-tangle ℓ (top) and its lifts $\ell/2$ (middle) and $\ell/3$ (bottom). Note that for illustration purposes the diagrams here are skewed and the three images have different horizontal scales. Thus the pictures don't reflect the fact that each n^{th} strand of each $\mathbb{Z}$-tangle starts at $(n, -1, 0)$ and ends at $(n, 1, 0)$.

to the k-fold cover $\mathbb{R}/(k\mathbb{Z}) \times I^2 \twoheadrightarrow \mathbb{R}/\mathbb{Z} \times I^2 = S^1 \times I^2$ and then reidentifying identifying $\mathbb{R}/(k\mathbb{Z})$ with S^1.

A similar construction is studied in [7]: In terms of annular 1-tangles, from some $\overline{\ell} \in \overline{A}$ we can create an unoriented knot (the *double branched cover* of $\overline{\ell}$) by taking the entire preimage of $\overline{\ell}$ in $\mathbb{R}/(2\mathbb{Z}) \times I^2$ and then embedding $\mathbb{R}/(2\mathbb{Z}) \times I^2$ into S^3 as in Section 2.2.

In the world of knotoids, lifting has established uses; for example, in [24] (Definition 15), the parity bracket polynomial of a knotoid could equivalently be defined as the bracket polynomial of the 2^{nd} lift of the knotoid.

Chapter 3

Heights and Sign Sequences

This chapter focuses on the signed heights of annular 1-tangles. The first goal of the chapter is Theorem 3.3.1.

3.1 Self-Intersecting Spanning Disks

The lemma in this section shows that loosening the definition of a spanning disk to allow certain self-intersections would not reduce the height of an annular 1-tangle or $\mathbb{Z}$-tangle.

Definition 3.1.1. A *generalized spanning disk* for an annular 1-tangle $\overline{\ell} \in \overline{A}$ is a piecewise-smooth map $\overline{d}\colon D^2 \to S^1 \times I^2$ with the following properties:

- The preimage of $\partial(S^1 \times I^2)$ under d is $\partial(D^2)$.
- The composite map
$$\partial D^2 \to \partial(S^1 \times I^2) = S^1 \times \partial(I^2) \twoheadrightarrow \partial(I^2),$$
where the first arrow is the restriction of $\overline{d}$ and last arrow is the projection onto the second factor, is a diffeomorphism.
- The intersections between (the images of) $\overline{d}$ and $\overline{\ell}$ are transverse and occur away from both the boundary and the self-intersection points of $\overline{d}$.
- The self-intersections of $\overline{d}$ avoid a neighborhood of the boundary.

Notice that intersections between $\overline{\ell}$ and $\overline{d}$ are signed, just as if $\overline{d}(D^2)$ were a typical (i.e. non-generalized) spanning disk: Using the orientatability of the domain D^2 and the orientation of the ambient $S^1 \times I^2$, there are two consistent ways to assign coorientations for all regular points in D^2 for the purposes of determining the signs of intersections with $\overline{\ell}$; we will use the assignment that, around the boundary, looks like it should for a spanning disk.

The definition of generalized spanning disks carries over to $\mathbb{Z}$-tangles as well, with $\mathbb{R} \times I^2$ in place of $S^2 \times I^2$.

Lemma 3.1.2. *If $\overline{d}$ is a generalized spanning disk for $\overline{\ell}$, then $\overline{d}$ has at least $\overline{h}(\overline{\ell})$ intersections with $\overline{\ell}$, at least $\overline{h}_+(\overline{\ell})$ of which are positive and at least $\overline{h}_-(\overline{\ell})$ of which are negative.*

The equivalent statement also holds for a $\mathbb{Z}$-tangle $\ell \in A$ and a generalized spanning disk $d \subset \mathbb{R} \times I^2$.

Proof. We will just prove the statement for an annular 1-tangle $\overline{\ell}$, since our proof carries over to the case of $\mathbb{Z}$-tangles without modification.

Given the generalized spanning disk $\overline{d}$, any small perturbation of $\overline{d}$ performed entirely away from the boundary and away from the intersections between $\overline{d}$ and $\overline{\ell}$ gives us a new generalized spanning disk with the same number of intersections of each sign with $\overline{\ell}$. Therefore, we can take $\overline{d}$ to be in general position, requiring that its self-intersections satisfy the hypotheses of Dehn's lemma as stated in [14] (page 193).

The version of Dehn's lemma that appears in the exposition in [14] (page 193) is slightly more specific in its stated conclusion than the original in [43]; this stronger version, applied to $\overline{d}$, asserts that there exists a resolution of the self-intersections of $\overline{d}$ such that the resulting surface has zero genus and hence consists only of a disk and possibly some 2-spheres. The resolution, of course, can be performed by modifying the image of $\overline{d}$ in just a small neighborhood of the singular set; since the intersections between $\overline{\ell}$ and $\overline{d}$ avoid the singular set, we can take the resolutions to not interfere with the intersections with $\overline{\ell}$. Let $\overline{D}$ be the disk component of such a resolution, which must have the same boundary as $\overline{d}$.

Now, every point of intersection between $\overline{D}$ and $\overline{\ell}$ also existed on $\overline{d}$, so the number of intersections between $\overline{d}$ and $\overline{\ell}$ is at least $\overline{H}(\overline{\ell}, \overline{D})$, which in turn is at least $\overline{h}(\overline{\ell})$ by the definition of $\overline{h}$. Note that the process of resolving self-intersections does not respect orientation, so a point of intersection between $\overline{D}$ and $\overline{\ell}$, considered as such, might not have the same sign as it does when considered as an intersection between $\overline{d}$ and $\overline{\ell}$. However, since $\overline{d}$ and $\overline{D}$ have the same boundary, they are homologous, which implies that the overall algebraic intersection numbers of $\overline{\ell}$ with each of $\overline{d}$ and $\overline{D}$ must be equal. Hence, the number of positive (resp. negative) intersections of $\overline{d}$ with $\overline{\ell}$ equals $H_\pm(\overline{\ell}, \overline{D})$, which is at least $\overline{h}_\pm(\overline{\ell})$. □

One immediate application of Lemma 3.1.2 is to the Birman–Hilden map:

Lemma 3.1.3. *For any $\overline{\ell} \in \overline{A}$, we have $h(s(\overline{\ell})) = \overline{h}(\overline{\ell})$. Furthermore, the two equations $h_\pm(s(\overline{\ell})) = \overline{h}_\pm(\overline{\ell})$ hold as well.*

Proof. Given $\overline{\ell} \in \overline{A}$, let $\ell = s(\overline{\ell})$.

Clearly we have $h(\ell) \leq \overline{h}(\overline{\ell})$ and $h_\pm(\ell) \leq \overline{h}_\pm(\overline{\ell})$: Any spanning disk $\overline{D}$ for $\overline{\ell}$ can be lifted to $\mathbb{R} \times I^2$ to create a spanning disk D for ℓ with $H_\pm(\ell, D) = \overline{H}_\pm(\overline{\ell}, \overline{D})$.

Conversely, suppose we are given a spanning disk D for ℓ. Then we can let $\overline{d}\colon D^2 \to S^1 \times I^2$ be the map obtained by composing an arbitrary identification $D^2 = D$ with ℓ and the covering map $\mathbb{R} \times I^2 \to S^1 \times I^2$. This $\overline{d}$ is an immersion, possibly with some self-intersections on its interior. The number of positive (resp. negative) intersections between $\overline{d}$ and $\overline{\ell}$ equals $H_+(\ell, D)$ (resp. H_-). The self-intersections of $\overline{d}$ correspond to intersections between D and translates of D by (nonzero) integers. By doing small perturbations, we can take D to be in general position with respect to integer translation. Under this assumption, the self-intersections of $\overline{d}$ will avoid $\overline{\ell}$ and therefore $\overline{d}$ will be a generalized spanning disk. Now we can apply Lemma 3.1.2 to the generalized spanning disk $\overline{d}$, obtaining $\overline{H}_\pm(\overline{\ell}, \overline{D}) = H_\pm(\ell, D)$. Therefore, $h(\ell) \geq \overline{h}(\overline{\ell})$ and $h_\pm(\ell) \geq \overline{h}_\pm(\overline{\ell})$. □

Therefore, on the level of isotopy classes, for $\overline{q}(\overline{\ell}) \in \overline{A}/\overline{\mathcal{I}}$ we have $h(\mathrm{Bh}(s_*(\overline{\ell}))) = \overline{h}(q(\overline{\ell}))$.

3.2 Disjoint Spanning Disks

When a pair of spanning disks for a single annular 1-tangle are disjoint from each other, we can observe restrictions on their heights:

Lemma 3.2.1. *If $\overline{D}_1$ and $\overline{D}_2$ are disjoint spanning disks for an annular 1-tangle $\overline{\ell} \in \overline{A}$ and if $\overline{H}(\overline{\ell}, \overline{D}_1) \geq \overline{H}(\overline{\ell}, \overline{D}_2)$, then $\overline{H}_\pm(\overline{\ell}, \overline{D}_1) \geq \overline{H}_\pm(\overline{\ell}, \overline{D}_2)$.*

Note that the equivalent statement for $\mathbb{Z}$-tangles does not hold.

Proof. Because $\overline{D}_1$ and $\overline{D}_2$ are disjoint, $\overline{D}_1 \cup \overline{D}_2$ partitions $S^1 \times I^2$ into two 3-balls. Let $\overline{\Sigma}$ denote $\overline{D}_1 \cup (-\overline{D}_2)$, that is, the oriented copy of $\overline{D}_1 \cup \overline{D}_2$ that uses the original orientation of $\overline{D}_1$ for that component and the opposite of the original orientation of $\overline{D}_2$ for the $\overline{D}_2$ component. Then $\overline{\Sigma}$ is nullhomologous as a 2-chain in $S^1 \times I^2$ rel boundary, the oriented boundary of one of the 3-balls.

The endpoints of $\overline{\ell}$, $(0, \pm 1, 0)$, are each either inside or outside of $\overline{\Sigma}$. The algebraic intersection number of $\overline{\ell}$ with $\overline{\Sigma}$ must be 0, 1, or -1. In symbols, we can write

$$-1 \leq \overline{H}_+(\overline{D}_1) - \overline{H}_-(\overline{D}_1) - \overline{H}_+(\overline{D}_2) + \overline{H}_-(\overline{D}_2) \leq 1. \tag{3.1}$$

(Here we have suppressed $\overline{\ell}$ in the notation.)

By assumption, we have

$$\overline{H}_+(\overline{D}_1) + \overline{H}_-(\overline{D}_1) \geq \overline{H}_+(\overline{D}_2) + \overline{H}_-(\overline{D}_2). \tag{3.2}$$

Combining (3.2) with each inequality in (3.1), we obtain $2\overline{H}_+(\overline{D}_1) - 2\overline{H}_+(\overline{D}_2) \geq -1$ and $2\overline{H}_-(\overline{D}_1) - 2\overline{H}_-(\overline{D}_2) \geq -1$. Therefore, $\overline{H}_\pm(\overline{D}_1) \geq \overline{H}_\pm(\overline{D}_2)$, as desired. □

The next fact provides opportunities to apply Lemma 3.2.1.

Lemma 3.2.2. *If $\overline{D}$ and $\overline{D}'$ are spanning disks for an annular 1-tangle $\overline{\ell}$ and $\overline{D}$ realizes the height of $\overline{\ell}$, then there is a sequence of spanning disks*

$$\overline{D}' = \overline{D}_0, \overline{D}_1, \ldots, \overline{D}_k = \overline{D}$$

such that consecutive disks are disjoint and the sequence $(\overline{H}(\overline{\ell}, \overline{D}_i))$ is nonincreasing in i.

Proof. We can choose a $\overline{D}_1$ that is just a copy of $\overline{D}'$ pushed slightly to one side and such that $\overline{D}$ and $\overline{D}_1$ intersect transversely. Then the intersections consist of a system $\mathcal{C}_1$ of disjoint circles and arcs ∂-properly embedded in both $\overline{D}$ and $\overline{D}_1$. We can also require in our choice of $\overline{D}_1$ that none of the intersection curves meet $\overline{\ell}$. For each $i \geq 1$, once we have chosen $\overline{D}_i$ we will form $\overline{D}_{i+1}$ in such a way that $\overline{D}_{i+1}$ has fewer total components of intersection with $\overline{D}$ than does $\overline{D}_i$. Let $\mathcal{C}_i$ be the system of intersections between $\overline{D}_i$ and $\overline{D}$.

Case 1: $\mathcal{C}_i$ is empty.

If $\mathcal{C}_i$ is empty, then $\overline{D}$ is disjoint with $\overline{D}_i$, so we set $k = i + 1$ and $\overline{D}_k = \overline{D}$. Because $\overline{D}$ realizes the height of $\overline{\ell}$, $\overline{H}(\overline{D}_i) \geq \overline{H}(\overline{D})$.

Case 2: $\mathcal{C}_i$ has an arc, but no circles.

If $\mathcal{C}_i$ contains an arc, but no circles, we can find an arc Γ that is innermost on $\overline{D}$. By innermost arc, we mean one for which we there is a disk E in $\overline{D}$ bounded by Γ such that E does not containing any arcs of $\mathcal{C}_i$ besides Γ. In particular, this means that $\partial E \setminus \Gamma$ is disjoint from $\overline{D}'$. Consider the arc of $\partial\overline{D}'$ that corresponds to the same portions of $\partial(I^2)$ as $\partial E \setminus \Gamma$. Then the disk E' in $\overline{D}_i$ bounded by this arc and Γ cannot contain any other arcs of $\mathcal{C}_i$, since their boundaries would have to be intersection points of $\partial E \setminus \Gamma$ with $\overline{D}'$. Therefore, Γ is also innermost in $\overline{D}_i$. In particular, $E \cap \overline{D}_i = E' \cap \overline{D} = \Gamma$.

Since $(\overline{D} - E) \cup E'$ is a spanning disk, its height is at least $\overline{H}(\overline{D})$. Therefore, E' has at least as many intersections with $\overline{\ell}$ as does E. Now let $\overline{D}_{i+1}$ be the result of slightly perturbing $(\overline{D}_i - E') \cup E$ to be disjoint from $\overline{D}_i$ and from E. Then $\overline{D}_{i+1}$ has fewer intersection curves than $\overline{D}_i$ with $\overline{D}$, and we have $\overline{H}(\overline{D}_{i+1}) \leq \overline{H}(\overline{D}_i)$.

Case 3: $\mathcal{C}_i$ contains a circle.

If there is at least one circle, there is an innermost circle C in $\overline{D}$. Then C bounds a disk $E \subset \overline{D}$ with $E \cap \overline{D}_i = C$. Let E' denote the disk in $\overline{D}_i$ bounded by C. In contrast with Case 2, C is not necessarily innermost in $\overline{D}_i$, so $E' \cap \overline{D}$ might be more than just C.

By Lemma 3.1.2, $\overline{\ell}$ has at least as many intersections with $(\overline{D} - E) \cup E'$ as with $\overline{D}$, so it intersects E' at least as many times as E. Therefore, we can proceed as in Case 2. Let $\overline{D}_{i+1} = (\overline{D}_i - E') \cup E$, and perturb it so that it is disjoint from D_i and E. Then $\overline{H}(\overline{D}_{i+1}) \leq \overline{H}(\overline{D}_i)$, and we have reduced the number of intersection curves with $\overline{D}$.

This covers all the cases, so we are done. □

Lemma 3.2.1 immediately implies that Lemma 3.2.2 can be strengthened as follows.

Lemma 3.2.3. *If $\overline{D}$ and $\overline{D}'$ are spanning disks for an annular 1-tangle $\overline{\ell}$ and $\overline{D}$ realizes the height of $\overline{\ell}$, then there is a sequence of spanning disks*

$$\overline{D}' = \overline{D}_0, \overline{D}_1, \ldots, \overline{D}_k = \overline{D}$$

such that consecutive disks are disjoint and the sequences $(\overline{H}_\pm(\overline{\ell}, \overline{D}_i))$ are both nonincreasing.

3.3 Additivity of Signed Heights

Theorem 3.3.1. *For each annular* 1*-tangle* $\overline{\ell}$, $\overline{h}(\overline{\ell}) = \overline{h}_+(\overline{\ell}) + \overline{h}_-(\overline{\ell})$.

Proof. Given an annular 1-tangle $\overline{\ell}$, a spanning disk $\overline{D}$ realizing the height, and any other spanning disk $\overline{D}'$, Lemma 3.2.3 implies that $\overline{D}$ has no greater positive or negative height than $\overline{D}'$. Therefore, $\overline{D}$ realizes the signed heights, and so we have

$$\overline{h}(\overline{\ell}) = \overline{H}(\overline{\ell}, \overline{D}) = \overline{H}_+(\overline{\ell}, \overline{D}) + \overline{H}_-(\overline{\ell}, \overline{D}) = \overline{h}_+(\overline{\ell}) + \overline{h}_-(\overline{\ell}).$$

□

Theorem 3.3.1 reduces questions about the height of an annular 1-tangle to questions about its signed heights, which form a *height pair* $(\overline{h}_+(\overline{\ell}), \overline{h}_-(\overline{\ell}))$. This will make it easier to compute the heights of some annular 1-tangles. The theorem also implies that all minimal attainable sequences for an annular 1-tangle are rearrangements of each other.

Theorem 3.3.1 contrasts with the case of crossing numbers of classical knots: Little ([37], page 774) conjectured that any two minimal diagrams of the same knot would have the same writhe (i.e. the number of positive crossings minus the number of negative ones). This conjecture would have followed from a stronger conjecture that signed crossing numbers determine the unsigned crossing number. However, there is a counterexample to both versions with crossing number 10 and signed crossing numbers 9 and 0.

Lemma 3.2.3 can also provide more detailed information about the set of attainable sign sequences for an annular 1-tangle. A *left shift move* of size k on a sign sequence S is the result of deleting k appearances of $(-,+)$ as a consecutive subsequence of S and then inserting k copies of $(+,-)$. Similarly, a *right shift move* deletes copies $(+,-)$ and adds copies of $(-,+)$. The deletions and insertions all happen at the same time. For example, a nontrivial shift move on $(-,-,+,+)$ must be a left shift of size 1, deleting the second and third entries. The possible results after inserting $(+,-)$ are $(+,-,-,+)$, $(-,+,-,+)$, and $(-,+,+,-)$. Right and left shift moves undo each other.

Lemma 3.3.2. *If* $\overline{D}_1$ *and* $\overline{D}_2$ *are disjoint spanning disks for an annular* 1*-tangle* $\overline{\ell}$ *and* $\overline{H}(\theta, \overline{D}_1) = \overline{H}(\overline{\ell}, \overline{D}_2)$*, then* $\operatorname{Seq}(\overline{\ell}, \overline{D}_1)$ *and* $\operatorname{Seq}(\overline{\ell}, \overline{D}_2)$ *differ by a shift move.*

Proof. Let $\overline{\Sigma}$ be the sphere $\overline{D}_1 \cup (-\overline{D}_2)$ as in the proof of Lemma 3.2.1. Without loss of generality (up to relabeling $\overline{D}_1$ and $\overline{D}_2$), we can assume that $(0,-1,0)$ is in the ball "outside" of $\overline{\Sigma}$. In the overall sequence of intersections of $\overline{\ell}$ with $\overline{\Sigma}$, the signs of the intersections alternate: The odd- and even-index intersections are negative and positive, respectively. Because $\overline{H}(\overline{\ell}, \overline{D}_1) = \overline{H}(\overline{\ell}, \overline{D}_2)$, the total number of intersections is even. Each odd-even pair of consecutive intersections has type $(+\overline{D}_2, +\overline{D}_1)$, $(-\overline{D}_1, -\overline{D}_2)$, $(-\overline{D}_1, +\overline{D}_1)$, or $(+\overline{D}_2, -\overline{D}_2)$. The sign sequences for $\overline{D}_1$ and $\overline{D}_2$ are obtained just by picking out the corresponding entries from the overall sequence of intersections with $\overline{\Sigma}$. Where we see $(+\overline{D}_2, +\overline{D}_1)$ or $(-\overline{D}_1, -\overline{D}_2)$ in the overall sequence, we will get a $+$ or a $-$, respectively, regardless of which spanning disk we consider. The other pairs will either show up as $(-,+)$ in $\operatorname{Seq}(\overline{\ell}, \overline{D}_1)$ only or as

$(+, -)$ in $\mathrm{Seq}(\bar{\ell}, \overline{D}_2)$ only. Therefore, $\mathrm{Seq}(\bar{\ell}, \overline{D}_2)$ is obtained from $\mathrm{Seq}(\bar{\ell}, \overline{D}_1)$ by a left shift move. □

Lemma 3.3.2 can be applied to annular 1-tangle invariants as follows.

Theorem 3.3.3. *If S and S' are two minimal attainable sequences for some $\bar{\ell} \in \overline{A}$, there is a sequence of minimal attainable sequences*

$$S = S_0, S_1 \dots, S_k = S'$$

such that each S_{i+1} differs from S_i by a shift move.

Proof. Suppose $\overline{D}$ and $\overline{D}'$ are both spanning disks realizing the height of $\bar{\ell}$. Then in the sequence $(\overline{D}_i)$ of spanning disks obtained from Lemma 3.2.3, each $\overline{D}_i$ realizes the height of $\bar{\ell}$. Therefore, applying Lemma 3.3.2 to the sequence $(\overline{D}_i)$ implies that $\mathrm{Seq}(\bar{\ell}, \overline{D})$ and $\mathrm{Seq}(\bar{\ell}, \overline{D}')$ are connected among minimal attainable sequences by shift moves. □

We will see later on how Theorem 3.3.3 can be used to distinguish distinct annular 1-tangles and knotoids.

3.4 Behavior of Attainable Sequences

Signed Heights under the Basic Involutions

The signed heights of annular 1-tangles behave in straightforward ways under the basic involutions.

Proposition 3.4.1. *For all $\bar{\ell} \in \overline{A}$, we have*

$$\bar{h}_\pm(\bar{\ell}) = \bar{h}_\pm(\mathrm{rev}(\bar{\ell})) = \bar{h}_\pm(\mathrm{mir}(\bar{\ell})) = \bar{h}_\mp(\mathrm{sym}(\bar{\ell})) = \bar{h}_\mp(\mathrm{rot}(\bar{\ell})).$$

More specifically, we can say the following.

Proposition 3.4.2. *If S is an attainable sign sequence for $\bar{\ell}$, then,*

1. $\mathrm{rev}(S)$ *is attainable for* $\mathrm{rev}(\bar{\ell})$,
2. S *is attainable for* $\mathrm{mir}(\bar{\ell})$,
3. $-S$ *is attainable for* $\mathrm{sym}(\bar{\ell})$, *and*
4. $-S$ *is attainable for* $\mathrm{rot}(\bar{\ell})$.

where $-S$ is the result of switching all terms $+ \leftrightarrow -$ in S and $\mathrm{rev}(S)$ *is the result of reversing the order.*

Proof. Given a spanning disk $\overline{D}$ for $\overline{\ell}$, any time we transform $\overline{\ell}$ by inverting any of the three coordinates x, y, z, we can do the same to $\overline{D}$ to get a new spanning disk for the new annular 1-tangle. Recall that hte orientation for $\overline{D}$ is always determined by how it sits in $S^1 \times I^2$. Flipping all z values in the whole picture does not change the sign of any intersection or the order they come in. Flipping all x values preserves the orientation of $\overline{\ell}$, but effectively changes the orientation of $\overline{D}$, so it flips the signs of all intersections. Flipping all y (and t) values essentially changes the orientation of $\overline{\ell}$ and also the order in which the intersections appear. These moves together generate the various basic involutions, so the behaviour of signed heights and sign sequences can easily be extrapolated. □

Proposition 3.4.1, together with Theorem 3.3.1, has implications for unsigned heights of annular 1-tangles, such as for rotatable annular 1-tangles, which are addressed in [7]. An annular 1-tangle $\overline{\ell}$ is *rotatable* if it is isotopic to $\mathrm{rot}(\overline{\ell})$.

Corollary 3.4.3. *Every rotatable $\overline{\ell} \in \overline{A}$ has even height.*

Multiplication and Concatenation

In this section we relate the set of attainable sign sequences for a product to the attainable sequences of its factors.

Proposition 3.4.4. *For any $\overline{\ell}_1$ and $\overline{\ell}_2$, if S_1 is an attainable sign sequence for $\overline{\ell}_1$ and S_2 is attainable for $\overline{\ell}_2$, then the concatenation S_1S_2 is attainable for $\overline{\ell}_1\overline{\ell}_2$.*

Proof. Given spanning disks $\overline{D}_1$ and $\overline{D}_2$ for $\overline{\ell}_1$ and $\overline{\ell}_2$, as noted in Section 2.4 we can assume $\overline{D}_i \cap (S^1 \times \{\pm 1\} \times I) = \{\frac{1}{2}\} \times \{\pm 1\} \times I$ for $i = 1, 2$. Then when we form the product $\overline{\ell}_1\overline{\ell}_2$, we can insert $\overline{D}_1$ and $\overline{\ell}_2$ into their corresponding places, and together they will form a "product spanning disk" for $\overline{\ell}_2\overline{\ell}_2$. The sign sequence for the product spanning disk is the concatenation of the sign sequences for the two spanning disks. □

Note that, as a special case of the statement above, if S is any attainable sequence for an annular 1-tangle $\overline{\ell}$, then the result of appending $+$ or $-$ to either end of S is also attainable for $\overline{\ell}$, because for any spanning disk $\overline{D}$ for $\overline{\ell}$ we can strech the part near $y = 1$ or $y = -1$ in either direction to create a new intersection with $\overline{\ell}$.

The next theorem is a converse for Proposition 3.4.4.

Theorem 3.4.5. *Any minimal attainable sequence for $\overline{\ell}_1\overline{\ell}_2$ is the concatenation of minimal attainable sequences for $\overline{\ell}_1$ and $\overline{\ell}_2$.*

We will prove Theorem 3.4.5 using a modification of original argument appearing in [49] for Corollary 3.4.7 below.

Corollary 3.4.6. *For two annular 1-tangles $\overline{\ell}_1$ and $\overline{\ell}_2$, $\overline{h}_\pm(\overline{\ell}_1\overline{\ell}_2) = \overline{h}_\pm(\overline{\ell}_1) + \overline{h}_\pm(\overline{\ell}_2)$.*

Corollary 3.4.7 (Turaev [49], Theorem 4.3)**.** *For two knotoids k_1 and k_2, the complexity/height of k_1k_2 equals the sums of the complexities/heights of k_1 and k_2.*

Lemma 3.4.8. *Suppose that $\overline{\ell}$ is an annular* 1*-tangle and $\overline{\Phi}$ is a compact oriented surface (not necessarily connected) ∂-propertly embedded in $S^1 \times I^2$ such that $\overline{\ell}$ intersects $\overline{\Phi}$ transversely, only one component of $\overline{\Phi}$ has nonempty boundary, and that component is a spanning disk whose orientation as a spanning disk matches its orientation as part of $\overline{\Phi}$. Then the sequence* $\mathrm{Seq}(\overline{\ell}, \overline{\Phi})$ *of signs of intersections of $\overline{\ell}$ with $\overline{\Phi}$ is attainable for $\overline{\ell}$.*

Proof. Let $\overline{D}$ be the disk component of $\overline{\Phi}$.

Case 1: Every closed component of $\overline{\Phi}$ is a sphere.

If $\overline{\Phi} = \overline{D}$, then of course $\mathrm{Seq}(\overline{\ell}, \overline{\Phi}) = \mathrm{Seq}(\overline{\ell}, \overline{D})$ is attainable.

If $\overline{\Phi}$ is not connected, some spherical component $\overline{\Phi}_1$ of $\overline{\Phi}$ must be "outermost" in the sense that no other sphere separates it from $\partial(S^1 \times I^2)$, and hence from $\overline{D}$. If the positive side of $\overline{\Phi}_1$ is the one facing outward, then we can choose an embedded path from $\overline{\Phi}_1$ to the positive side of $\overline{D}$ such that the path does not otherwise intersect $\overline{\Phi}$ or $\overline{\ell}$. Then we can incorporate $\overline{\Phi}_1$ into $\overline{D}$ in an orientation-respecting way by adding an annulus to connect $\overline{\Phi}_1$ to $\overline{D}$ and deleting disks in $\overline{D}$ and $\overline{\Phi}_1$ around the path's endpoints. If instead the negative side of $\overline{\Phi}_1$ was facing outward, we can do the same but with the negative side of $\overline{D}$. Doing this several times replaces $\overline{\Phi}$ with a spanning disk and realizes the sign sequence as attainable.

Case 2: General Case.

Each closed component of $\overline{\Phi}$ is separating in $S^1 \times I^2$, and as in Case 1 each component has a clear inside and outside based on $\partial(S^1 \times I^2)$, regardless of their orientations. Call a spherical component of $\overline{\Phi}$ *trivial* if it intersects $\overline{\ell}$ twice and if $\overline{\ell}$ is unknotted inside the sphere. Let $\mathcal{F}$ be the set of closed components that are not trivial spheres. If $\mathcal{F}$ is nonempty, consider an innermost element $\overline{\Phi}_1$ of $\mathcal{F}$. Inside of $\overline{\Phi}_1$ are some number of segments of $\overline{\ell}$. Some of these segments could have trivial spheres attached. Let us delete $\overline{\Phi}_1$ and replace it with several trivial spheres: One sphere is added for each segment of $\overline{\ell}$ inside $\overline{\Phi}_1$, surrounding the segment and all preexisting trivial spheres on that segment. The new spheres can be oriented appropriately so that we have not changed $\mathrm{Seq}(\overline{\ell}, \overline{\Phi})$. Repeating this process renders $\mathcal{F}$ empty and reduces us to Case 1. □

Lemma 3.4.9. *For any annular* 1*-tangle $\overline{\ell}_1$ and a knot-type $\overline{\ell}_0$ representing $\kappa^\bullet$, a sign sequence is attainable for $\overline{\ell}_0\overline{\ell}_1$ if and only if it is attainable for $\overline{\ell}_1$.*

Proof. For convenience, let us write $\overline{\ell} = \overline{\ell}_0\overline{\ell}_1$. It is immediate that any attainable sequence for $\overline{\ell}_1$ is attainable for $\overline{\ell}$.

For the other direction, suppose we have a spanning disk $\overline{D}$ for $\overline{\ell}$. Pick a ball Ω in the interior of $S^1 \times I^2$ such that $\partial\Omega$ intersects $\overline{D}$ transversely and the restriction of $\overline{\ell}$ to Ω is κ in the form of a 1-tangle. If $\overline{\ell}$ intersects $\overline{D}$ inside Ω, we can push these intersections to the outside: Choose a subinterval of $\overline{\ell} \cap \Omega$ containing the intersections with $\overline{D}$ as well as one of the two endpoints, then delete from Ω a regular neighborhood of that interval. Therefore we can choose Ω so that $\overline{\ell} \cap \overline{D} \cap \Omega$ is empty.

Let v_0 and v_1 be the intersections of $\bar{\ell}$ with $\partial\Omega$, assigned such that $\bar{\ell}$ is oriented from v_0 to v_1. Each component of $\overline{D} \cap \Omega$ is a genus-0 surface (with boundary) properly embedded in Ω, and each component of $\overline{D} \cap \partial\Omega$ is an oriented circle with winding number 0, 1, or -1 around $\partial\Omega - \{v_0, v_1\}$. Suppose there is at least one circle with winding number 0. Then there is an innermost such circle. We can cut $\overline{D}$ along this circle and fill in two disks on either side of $\partial\Omega$ to obtain a new spanning surface consisting of a disk and a sphere. Doing these repeatedly, we obtain a surface $\overline{\Phi}$ as in Lemma 3.4.8 (specifically, as in Case 1) such that $\mathrm{Seq}(\bar{\ell}, \overline{\Phi}) = \mathrm{Seq}(\bar{\ell}, \overline{D})$ and and such that every component of $\overline{\Phi} \cap \partial\Omega$ has winding number ± 1. There might now be some spherical components of $\overline{\Phi}$ contained entirely within Ω, but they do not intersect $\bar{\ell}$.

Now, let us label the components of $\overline{\Phi} \cap \Omega$ as $E_1, \ldots, E_n$. Each separates Ω into two regions, and each is disjoint from the others and from $\bar{\ell}$. The components of $\overline{\Phi} \cap \partial\Omega$, all concentric circles, have an order based on how they are arranged from v_0 to v_1 and so can be indexed $1, \ldots, m$. For each i, let $\mathrm{wn}(i)$ be the winding number of the i^{th} circle and let $c(i)$ be the index of its component in $\overline{\Phi} \cap \Omega$. For each $j \in \{1, \ldots, n\}$, the sum of the $\mathrm{wn}(i)$ over all i with $c(i) = j$ must be 0, because $\bar{\ell}$ does not intersect E_j. For similar reasons, for all j and all i_1 and i_2 with $i_1 < i_2$ and $c(i_1) = c(i_2) \neq j$, the sum of the $\mathrm{wn}(i)$ over all i with $c(i) = j$ and $i_1 < i < i_2$ is also 0.

Now, to show that $\mathrm{Seq}(\bar{\ell}, \overline{\Phi})$ is attainable for $\bar{\ell}_1$, we will create a spanning of $\bar{\ell}_1$ by deleting and replacing the interior of Ω. Let Ω' be a standard 3-ball, and choose an orientation-respecting identification $\partial\Omega' \cong \partial\Omega$. Let λ be an unknotted strand properly embedded in Ω' from v_0 to v_1. Now consider a partition of the components of $\overline{\Phi} \cap \partial\Omega$ into pairs such that (a) paired components have opposite winding numbers, (b) paired components come from the same component of $\overline{\Phi} \cap \Omega$, and (c) for $i_1 < i_2 < i_3 < i_4$, we do not have i_1 paired to i_3 and i_2 to i_4. (A simple induction argument shows this is possible.) Now, we connect each pair of components with an unknotted annulus disjoint from λ and disjoint from the other annuli. See Figure 3.1.

Now we glue Ω' along $\partial\Omega$ to the closure of the complement of Ω to obtain a spanning surface $\overline{\Phi}'$ for $\bar{\ell}_1$, with $\mathrm{Seq}(\bar{\ell}_1, \overline{\Phi}') = \mathrm{Seq}(\bar{\ell}, \overline{D})$. Because we replaced each component of $\overline{\Phi} \cap \Omega$ with several annuli, we have not created any higher-genus components by replacing $\overline{\Phi}$ with $\overline{\Phi}'$. Therefore $\overline{\Phi}'$ is a union of a disk with spheres, and so by Lemma 3.4.8, $\mathrm{Seq}(\bar{\ell}, \overline{D})$ is attainable for $\bar{\ell}_1$. $\square$

Proof of Theorem 3.4.5. Given a spanning disk $\overline{D}$ for a product $\bar{\ell} = \bar{\ell}_1 \bar{\ell}_2$ realizing its height, we wish for there to be decomposing annulus $\overline{Q}$ expressing $\bar{\ell}$ as $\bar{\ell}_1 \bar{\ell}_2$ and also expressing $\overline{D}$ as a product spanning disk. By Lemma 3.4.9, we may assume that $\bar{\ell}_1$ and $\bar{\ell}_2$ each have no knot-type factors. (Without this assumption, the desired decomposing annulus would not necessarily exist.)

By construction, there is a decomposing annulus $\overline{Q}$ coming from $\bar{\ell}$ as a product of $\bar{\ell}_1$ and $\bar{\ell}_2$. Necessarily, the two vertices of $\bar{\ell}$ lie on opposite sides of $\overline{Q}$, and $\bar{\ell}$ intersects $\overline{Q}$ once transversely. We can choose $\overline{Q}$ such that $\overline{D}$ intersects $\overline{Q}$ transversely as well. Then the intersection of $\overline{D}$ and $\overline{Q}$ consists of a line segment and possibly several circles. If the number

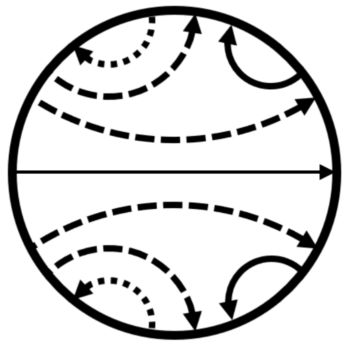

Figure 3.1: A cross section of Ω'. We can obtain Ω' by rotating around the center axis, which represents $\overline{\ell}'$. In this example, $\overline{\Phi} \cap \Omega$ had three components, indicated by the three types of curved arrow.

of circles is 0, then $\overline{Q}$ cuts $\overline{D}$ into two disks which are spanning disks for $\overline{\ell}_1$ and $\overline{\ell}_2$, so we are done.

If there are some circles, we can pick one which is innermost in $\overline{D}$. This bounds a disk E in $\overline{D}$ which does not otherwise intersect $\overline{D}$ or $\overline{Q}$. The circle is also nullhomologous on $\overline{Q}$ and separates $\overline{Q}$ into a pair of pants $\overline{Q}_1$ and a disk $\overline{Q}_2$. Let Ω denote the ball in $S^1 \times I^2$ with boundary $\overline{Q}_2 \cup E$.

The intersection of $\overline{\ell}$ with $\overline{Q}$ could be on either $\overline{Q}_1$ or $\overline{Q}_2$, but regardless, $\overline{\ell}$ cannot intersect E more times than it intersects $\overline{Q}_2$, by Lemma 3.1.2. Therefore, $\overline{\ell}$ either intersects $\overline{Q}_1$ once without intersecting $\overline{Q}_2$ or E, or it intersects $\overline{Q}_2$ and E once each but not $\overline{Q}_1$. In the latter case there is a 1-tangle inside of Ω, but by our assumption of no knot-type factors, the tangle is unknotted.

Let $\overline{Q}'$ be the sphere formed by pushing $\overline{Q}_2$ through Ω and past E, so $\overline{Q}'$ is a slight perturbation of $\overline{Q}_1 \cup E$ and there are fewer circular intersections of $\overline{D}$ with $\overline{Q}'$ than with $\overline{Q}$. Since Ω either does not intersect $\overline{\ell}$ or contains only an unknotted segment between $\overline{Q}_2$ and E, $(\overline{\ell}, \overline{Q}')$ is isotopic to $(\overline{\ell}, \overline{Q})$, so $\overline{Q}'$ still decomposes $\overline{\ell}$ as $\overline{\ell}_1\overline{\ell}_2$.

Repeating the above steps yields an annulus intersecting $\overline{\ell}$ once and $\overline{D}$ in only an interval, so it decomposes $(\overline{\ell}, \overline{D})$ as a product of $(\overline{\ell}_1, \overline{D}_1)$ and $(\overline{\ell}_2, \overline{D}_2)$, as desired. Then $\mathrm{Seq}(\overline{\ell}, \overline{D})$ is the concatenation of $\mathrm{Seq}(\overline{\ell}_1, \overline{D}_1)$ with $\mathrm{Seq}(\overline{\ell}_2, \overline{D}_2)$. Since $\mathrm{Seq}(\overline{\ell}_1\overline{\ell}_2, \overline{D})$ is minimal, each $\mathrm{Seq}(\overline{\ell}_i, \overline{D}_i)$ is also minimal. □

Signed Heights under Lifting

It is straightforward, but slightly complicated, to compute how attainable sign sequences behave with respect to lifting: For an annular 1-tangle $\overline{\ell}$ and spanning disk $\overline{D}$, there are k lifts of $\overline{D}$ to a shortcut for $\overline{\ell}/k$. The total number of positive (resp. negative) intersections of $\overline{\ell}/k$ with all such lifts is equal to $\overline{H}_+(\overline{\ell}, \overline{D})$ (resp. $\overline{H}_-(\overline{\ell}, \overline{D})$). Of course, that amount must be at least k times the minimal number of positive/negative intersections with each of the k lifts of $\overline{D}$. Therefore we have the result below.

Proposition 3.4.10. *For all $\overline{\ell} \in \overline{A}$ and k, $k\overline{h}_{\pm}(\overline{\ell}/k) \leq \overline{h}_{\pm}(\overline{\ell})$.*

Furthermore, we can obtain attainable sign sequences for $\overline{\ell}/k$ from attainable sequences for $\overline{\ell}$ in the following way. Given a sign sequence S of length H, for each $i \in \{0, \dots, H\}$ let $\mathrm{PS}_S(i)$ be the sum of the terms of S from indices 1 to i. For $i \in \{1, \dots, H\}$, let $\mathrm{PSM}_S(i)$ be the maximum of $\mathrm{PS}_S(i-1)$ and $\mathrm{PS}_S(i)$. Then for $m \in \mathbb{Z}/k\mathbb{Z}$, let $(S)^m$ be the subsequence of S consisting of only the terms from indices i with $\mathrm{PSM}_S(i) \equiv m \pmod k$. Given $\overline{\ell}$ and $\overline{D}$, we can label the k lifts of $\overline{D}$ as $\overline{D}^1, \dots, \overline{D}^n$ in such a way that they increment left in $S^1 \times I^2$, and $(0, -1, 0)$ sits between $\overline{D}^n$ and $\overline{D}^1$. Then the i^{th} intersection of $\overline{\ell}$ with $\overline{D}$ lifts to an intersection of $\overline{\ell}/k$ with $\overline{D}^{\mathrm{PSM}(i)}$, so for each m, $\mathrm{Seq}(\overline{\ell}/k, \overline{D}^m) = \mathrm{Seq}(\overline{\ell}, \overline{D})^m$. This implies the following.

Proposition 3.4.11. *For every attainable sequence S for $\overline{\ell}$, each $(S)^m$ is attainable for $\overline{\ell}/k$.*

3.5 Bounds on Signed Height

This section addresses lower bounds for signed heights, which of course are typically more difficult to find than upper bounds.

In this section we will usually use knotoids instead of annular 1-tangles in order to be able to use crossings.

The signed heights of a knotoid can be bounded individually by some polynomial invariants, in particular the Turaev polynomial $\langle\langle k \rangle\rangle_\circ$ of [49] (Section 8.1) and the index polynomial F_k of [30] (Definition 3.1). For a nonzero Laurent polynomial $\pi(t) \in \mathbb{Z}[t^{\pm 1}]$, we will write $\deg^+(\pi)$ for $\max\{\mathrm{maxdeg}(\pi), 0\}$ and $\deg^-(\pi)$ for $\max\{-\mathrm{mindeg}(\pi), 0\}$. We also set $\deg^{\pm}(0) = 0$. For a Laurent polynomial in multiple variables, the signed degree in a specific variable will be denoted by (for example) $\deg_t^{\pm}$.

Writhes

Given a crossing c in a knotoid diagram K, there is a unique resolution of c that respects orientation. This resolution creates an oriented diagram with two components, a loop K_0 and an interval K_1 with the same endpoints as K. The winding number of K_0 around the twice-punctured sphere is called the *intersection index* of c, denoted $\mathrm{Ind}(c)$ ([30], page 3, under the notation i). The index is equal to the intersection number of K_0 with any shortcut, or with K_1. Note that the index of a crossing doesn't depend on any over/under information. If a crossing has index n, it will be called an *n-crossing*.

Definition 3.5.1. For nonzero n, the *n-writhe* $J_n(K)$ of K is half the sum of the signs of the n-crossings.

Theorem 3.5.2 (Kim–Im–Lee [30], Theorem 2.4)**.** *For nonzero n, the n-writhe is a knotoid invariant.*

Remark. Our convention differs from [30] (Definition 2.2) by a factor of 2; they omit the word "half" in Definition 3.5.1. Under our convention, the n-writhe is still an integer: Any knotoid diagram can be turned into a diagram for the trivial knotoid by switching the signs of crossings such that each "late" strand passes over each "early" strand. Each such switch changes the n-writhe by an integer, and the n-writhe of the trivial knotoid is 0, so all n-writhes of all knotoids are integers. However, what we say here does not apply in general to virtual knotoids, which are considered in [30] alongside classical knotoids.

The following are immediate consequences of the definition of n-writhe:

Proposition 3.5.3. *For a knotoid k, we have the following:*

1. $J_n(\mathrm{rev}(k)) = J_n(k)$
2. $J_n(\mathrm{mir}(k)) = -J_n(k)$
3. $J_n(\mathrm{sym}(k)) = -J_{-n}(k)$
4. $J_n(\mathrm{rot}(k)) = J_{-n}(k)$

Proposition 3.5.4. *For knotoids k_1 and k_2, $J_n(k_1k_2) = J_n(k_1) + J_n(k_2)$.*

The n-writhes of a knotoid can be encoded in the coefficients of a polynomial. The *index polynomial* ([30], Definition 3.1, under the notation F) for k is

$$\mathrm{IP}_k(t) = \sum_{n \neq 0} J_n(k)(t^n - 1) \in \mathbb{Z}[t, t^{-1}].$$

This is closely related to its similarly-named predecessor, the *affine index polynomial* of [23] (Definition 12, under the notation P), defined by

$$\mathrm{AIP}_k(t) = \sum_c \mathrm{sign}(c)(t^{\mathrm{wt}(c)} - 1),$$

where $\mathrm{wt}(c)$ is $\mathrm{sign}(c)\,\mathrm{ssgn}(c)\,\mathrm{Ind}(c)$, and $\mathrm{ssgn}(c)$ is as shown in Figure 3.2. Note that $\mathrm{wt}(c)$ differs from $\mathrm{Ind}(c)$ only by sign. The affine index polynomial satisfies $\mathrm{AIP}_k(t) = \mathrm{AIP}_k(t^{-1})$ for all k ([23], Theorem 4.10), so it is related to the index polynomial by the formula

$$\mathrm{AIP}_k(t) = \mathrm{IP}_k(t) + \mathrm{IP}_k(t^{-1}). \tag{3.3}$$

The degree of the affine index polynomial was shown to be a lower bound for the height of a knotoid in [23] (Theorem 4.12). Because of the relationship in eq. (3.3), this is equivalent to Proposition 3.12 of [30]. Proposition 3.5.6, together with Theorem 3.3.1, is an improvement on this bound in the case that $\deg^+(\mathrm{IP}_k)$ and $\deg^-(\mathrm{IP}_k)$ are both positive.

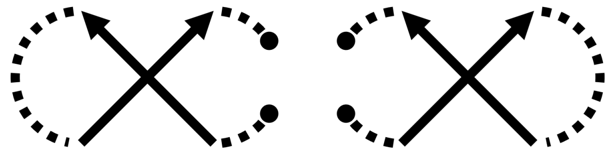

Figure 3.2: Sequential signs of crossings.

Proposition 3.5.5. *Any attainable sequence for k must have a consecutive subsequence adding up to* $\deg^+(\mathrm{IP}_k)$*, and a consecutive subsequence adding up to* $-\deg^-(\mathrm{IP}_k)$*.*

In the theorem above, of course, we treat $+$ terms as $+1$ and $-$ as -1.

Proof. For nonzero n, if $J_n(k) \neq 0$, any diagram K for k must have an n-crossing c. Then given any shortcut a for K, the segment of K starting and ending at c has, algebraically, n intersections with a, so the signs in the corresponding segment of $\mathrm{Seq}(K,a)$ add up to n. □

Proposition 3.5.6. *For a knotoid k, $\overline{h}_\pm(k) \geq \deg^\pm(\mathrm{IP}_k)$.*

Proof. If $J_n(k) \neq 0$, then as above, in any attainable sign sequence for k there is a consecutive subsequence with sum n. Therefore, for positive n there must be at least n appearances of $+$, and for negative n there are at least $-n$ appearances of $-$. This proves that the positive/negative height of k is bounded below by the positive/negative degree of IP_k. □

Theorem 3.5.7. *If k is a knotoid such that the bounds in Proposition 3.5.6 are equalities, then k has a unique minimal attainable sign sequence.*

Proof. Suppose that $\overline{h}_\pm(k) = \deg^\pm(\mathrm{IP}_k)$. A minimal attainable sign sequence contains $\overline{h}_+(k)$ copies of $+$ and $\overline{h}_-(k)$ copies of $-$, and by Proposition 3.5.5, the terms of the same sign must all be consecutive. Therefore, any minimal attainable sign sequence is one of $(+,\ldots,+,-,\ldots,-)$ or $(-,\ldots,-,+,\ldots,+)$. Call these two sequences S_1 and S_2, respectively. To show that only one of these can be attainable, we consider several cases.

Case 1: $\overline{h}_+(k)$ or $\overline{h}_-(k)$ is 0.

If one of the signed heights is zero, then all of the terms are the same sign, and $S_1 = S_2$.

Case 2: $\overline{h}_+(k), \overline{h}_-(k) > 1$.

In this case, S_1 and S_2 are not related by a shift move. There are no other minimal attainable sequences, so by Theorem 3.3.3, they cannot both be attainable.

Case 3: $\overline{h}_+(k)$ or $\overline{h}_-(k)$ is 1 and neither is 0.

Supposing that S_1 and S_2 are both attainable, they are the only minimal attainable sequences. By Lemma 3.2.3, there are disjoint spanning disks $\overline{D}_1$ and $\overline{D}_2$ for a $\overline{\ell} \in \overline{A}$ corresponding to k such that $\mathrm{Seq}(\overline{\ell}, \overline{D}_i) = S_i$ for $i = 1, 2$.

Since $\overline{D}_1$ and $\overline{D}_2$ are both minimal, by Theorem 3.3.1 they correspond to the same spanning disk framing for $\overline{\ell}$. Together they bound a thickened disk Ω in $S^1 \times I^2$ that does not include $(0, \pm 1, 0)$. There are $\overline{h}(\overline{\ell})$ segments of $\overline{\ell}$ in Ω, one entering and leaving by $\overline{D}_1$, one entering and leaving by $\overline{D}_2$, and the rest crossing from one side to the other. Without loss of generality, we can assume that $\overline{\ell}$ intersects $\overline{D}_2$ before $\overline{D}_1$. Then $\overline{h}_-(\overline{\ell})$ must be 1, and the intersections come in the order

$$(-\overline{D}_2, +\overline{D}_2, \ldots, +\overline{D}_1, -\overline{D}_1),$$

where the "…" consists of $\overline{h}_+(\overline{\ell}) - 1$ consecutive copies of $(+\overline{D}_1, +\overline{D}_2)$.

By isotopy (possibly moving the boundary of $S^1 \times I^2$, but leaving $(0, \pm 1, 0)$ fixed throughout and without changing the sign sequences), we can assume both $\overline{D}_1$ and $\overline{D}_2$ are simultaneously planes in S^1, both parallel to the z direction ([27], point (9) in the appendix). By then adjusting $\overline{\ell}$ to be in general position with respect to the projection to $S^1 \times I$, we can obtain a diagram K for k with two shortcuts a_1 and a_2 corresponding to $\overline{D}_1$ and $\overline{D}_2$ that are disjoint on their interiors.

Since any crossing c of K lies either in the region of S^2 corresponding to Ω or in its complement, the loop on K from c to c has an even total number of intersections with a_1 and a_2. Therefore, if that loop includes the negative intersection with a_1, it also includes at least one positive a_1 intersection, so if we measure the index of c by intersections of the loop with a_1, the index is nonnegative.

Since all crossings have nonnegative index, all negative writhes are 0, contradicting the assumption that $\overline{h}_-(k) = \deg^-(\mathrm{IP}_k)$. This proves Theorem 3.5.7. □

The Turaev Polynomial

A *state* of a diagram K is a function from the set of crossings to $\{-, +\}$. For each state ξ, the ξ-smoothing of K is given by smoothing each crossing according to Figure 3.3. The sum of $\xi(c)$ over all crossings is denoted $\mathrm{tot}(\xi)$, and the number of embedded circles in the diagram after smoothing by ξ is $\mathrm{circ}(\xi)$. (There is also one embedded interval, which is not counted.) The *bracket polynomial* of K ([49], Section 7.1) is then

$$\langle K \rangle = \sum_s \alpha^{\mathrm{tot}(\xi)} (-\alpha^2 - \alpha^{-2})^{\mathrm{circ}(\xi)} \in \mathbb{Z}[\alpha, \alpha^{-1}].$$

The bracket polynomial is invariant under Reidemeister moves I', II, and III, so it is a "framed knotoid" invariant. A Reidemeister I move changes the bracket polynomial by a factor of $-\alpha^{-3}$, so the *normalized bracket polynomial* defined by

$$\langle K \rangle_\circ = (-\alpha)^{-3\,\mathrm{wr}(K)} \langle K \rangle$$

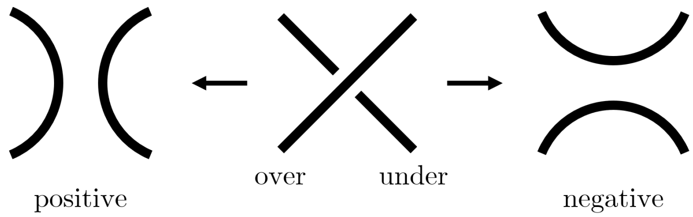

Figure 3.3: Smoothings for the bracket polynomial.

in [49] is an unframed invariant.

There is also a two-variable version of the bracket polynomial, called the *extended bracket polynomial* in [49] (Section 8.1) or the *Turaev polynomial* as in [34] (Section 2.2). For a diagram K and shortcut a, let $a(K)$ denote the algebraic height

$$\overline{H}_+(K,a) - \overline{H}_-(K,a),$$

and for any state, let $a(\xi)$ be the algebraic height of the interval component of the ξ-smoothing of K, with its natural orientation. Then the Turaev polynomial of (K,a) is

$$\langle\langle K,a\rangle\rangle = \sum_s \alpha^{\mathrm{tot}(\xi)} u^{a(\xi)} (-\alpha^2 - \alpha^{-2})^{\mathrm{circ}(\xi)} \in \mathbb{Z}[\alpha^{\pm 1}, u^{\pm 1}].$$

The normalized version

$$\langle\langle K,a\rangle\rangle_\circ = (-\alpha)^{-3\,\mathrm{wr}(K)} u^{-a(K)} \langle\langle K,a\rangle\rangle$$

is a knotoid invariant and always takes values in $\mathbb{Z}[\alpha^{\pm 2}, u^{\pm 2}]$.

The height of a knotoid satisfies $2\overline{h}(k) \geq \deg_u^+(\langle\langle k\rangle\rangle_\circ) + \deg_u^-(\langle\langle k\rangle\rangle_\circ)$ ([49], Section 8.3). Proposition 3.5.8 below does not improve this bound on the overall height but is the equivalent statement for the signed heights.

Proposition 3.5.8. *For a knotoid k, $2\overline{h}_\pm(k) \geq \deg_u^\mp(\langle\langle k\rangle\rangle_\circ)$.*

Proof. Fix a diagram K representing a knotoid k and a shortcut a for K. For any state ξ, the ξ-smoothing of K only has as many intersections with a as K does. In particular, the interval component of the smoothing has no more than

$$\overline{H}_+(K,a) + \overline{H}_-(K,a)$$

positive or negative intersections with a, so we have

$$-2\overline{H}_+(K,a) \leq a(\xi) - a(K) \leq 2\overline{H}_-(K,a).$$

Therefore, the u exponents of $\langle\langle k\rangle\rangle_\circ$ are no more than $2\overline{h}_-(k)$ and no less than $-2\overline{h}_+(k)$. □

A categorification of the Turaev polynomial, the triply-graded *winding homology*, is defined in [34]. The corresponding Poincaré polynomial is denoted $W_k(t,\alpha,u)$ and satisfies

$$W_k(-1,\alpha,u) = \langle\langle k\rangle\rangle_\circ$$

for every k. The winding homology is the homology of a chain complex in which each generator is given a u-grading $a(\xi)-a(K)$ for some state ξ, so in addition to Proposition 3.5.8 we can also say that

$$2\overline{h}_\pm(k) \geq \deg_u^\mp(W_k(t,\alpha,u)).$$

3.6 Low Heights

As we saw before, annular 1-tangles with height 0 are, up to isotopy, naturally in one-to-one correspondence with oriented knots. We can also make some general statements about annular 1-tangles with height 1 or 2.

Height One

Lemma 3.2.3 allows us to characterize annular 1-tangles of height 1 using 2-tangles. Suppose we are given a 2-tangle T in the 3-ball $[-\frac{1}{2},\frac{1}{2}]\times I^2$ connecting $(0,-1,0)$ to $(\frac{1}{2},0,0)$ and $(-\frac{1}{2},0,0)$ to $(0,1,0)$. Then we can form an annular 1-tangle $\overline{\ell}_T$ by placing $[-1\frac{1}{2},\frac{1}{2}]\times I^2$ into $S^1\times I^2$ in the natural way, effectively gluing the two sides $\{\pm\frac{1}{2}\}\times I^2$. This annular 1-tangle clearly has $\overline{h}_+(\overline{\ell}_T)\leq 1$ and $\overline{h}_-(\overline{\ell}_T)=0$. Let $\mathcal{T}$ be the set of isotopy classes of such tangles T that also satisfy the requirements (a) no ball inside $[-\frac{1}{2},\frac{1}{2}]\times I^2$ contains a nontrivial 1-tangle and (b) T is not isotopic to the trivial tangle formed by two straight line segments.

An annular 1-tangle of height 1 is prime if and only if it has no knot-type factor. Height-1 annular 1-tangles each have height pair $(1,0)$ or $(0,1)$, and the two types are in bijective correspondence via rotation.

Theorem 3.6.1. *The map* $[T]\mapsto\overline{q}(\overline{\ell}_T)$ *is a bijection from* $\mathcal{T}$ *to the set of prime elements of* $\overline{A}/\overline{\mathcal{I}}$ *that have height pair* $(1,0)$.

Proof. Given $[T]\in\mathcal{T}$, it is immediate that $\overline{\ell}_T$ is an element of $\overline{A}$ with $h_+(\overline{\ell}_T)\leq 1$ and $h_-(\overline{\ell}_T)=0$. It is also clear that the isotopy class of $\overline{\ell}_T$ depends only on the isotopy class of T. We have to show that $h_+(\overline{\ell}_T)=1$, that the class of $\overline{\ell}_T$ is prime, and that the map is both injective and surjective.

For such a $[T]\in\mathcal{T}$, there is a spanning disk D_0 for $\overline{\ell}_T$ such that $\mathrm{Seq}(\overline{\ell}_T,D_0)=(+)$ and such that T can be recovered by cutting along D_0. Suppose D_1 is another spanning disk for $\overline{\ell}_T$ such that D_0 and D_1 are disjoint and D_1 also has sign sequence $(+)$. Then D_0 and D_1 together divide $S^1\times I^2$ into two regions, and inside the one that doesn't contain the endpoints is a 1-tangle that, by condition (a) of the definition of $\mathcal{T}$, is unknotted. Therefore, cutting $\overline{\ell}_T$ along D_1 gives back a 2-tangle isotopic to T.

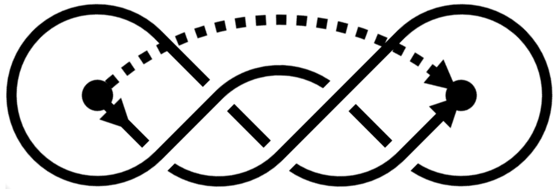

Figure 3.4: The Kinoshita knotoid.

Suppose $\overline{\ell}_T$ has $h_+ = 0$. Then by Lemma 3.2.3 and the previous paragraph, there is a spanning disk D disjoint from D_0 with $\overline{H}(\overline{\ell}_T, D) = 0$. Then $D_0 \cup D$ splits $\overline{\ell}_T$ into two 1-tangles, which must both be trivial, contradicting condition (b) of the definition of $\mathcal{T}$. Therefore, $\overline{\ell}_T$ has height 1.

Now suppose that $\overline{q}(\overline{\ell}_T)$ is not prime. Then by additivity of heights, it has a nontrivial knot-type factor. Therefore, there is a ball in $S^1 \times I^2$ intersecting $\overline{\ell}_T$ in a single strand making a prime knot κ in the ball. Then as in the proof of Theorem 2.10 in [35], the ball can be chosen to avoid $\{\frac{1}{2}\} \times I^2$. (See the proof of Lemma 5.4.1 below for details.) This contradicts the definition of $\mathcal{T}$, so we can conclude that $\overline{q}(\overline{\ell}_T)$ must actually be prime.

Knowing that $\overline{\ell}_T$ has height 1, Lemma 3.2.3 now implies that no other T' with $[T'] \neq [T]$ has $\ell_{T'} = \overline{\ell}_T$, so the map is injective.

For any prime annular 1-tangle $\overline{\ell}$ with positive and negative heights 1 and 0, we can obtain a $[T] \in \mathcal{T}$ with $\overline{\ell}_T = \overline{\ell}$ by choosing a spanning disk D for $\overline{\ell}$ with height 1 and cutting. Since $\overline{\ell}$ is prime and has height 1, T satisfies conditions (a) and (b). □

Height Two

Consider the following two examples of annular 1-tangles with height 2.

Example 3.6.2. The Kinoshita knotoid, shown in Figure 3.4, is an example of a nontrivial knotoid with trivial overpass and underpass closures (see [31], Figure 1, for the corresponding theta-curve). The knotoid diagram shown has a shortcut with sign sequence $(-,+)$, and the index polynomial is $t^{-1} - 2 + t$. Therefore, if $\overline{\ell}$ is the corresponding annular 1-tangle, then we have $\overline{h}_+(\overline{\ell}) = \overline{h}_-(\overline{\ell}) = 1$ and, by Theorem 3.5.7, $(-,+)$ is the only minimal attainable sign sequence for $\overline{\ell}$.

The annular 1-tangle $\overline{\ell}$ satisfies $\mathrm{rev}(\overline{\ell}) = \mathrm{rot}(\overline{\ell})$. Note that neither the index polynomial nor the Turaev polynomial distinguishes $\mathrm{rot}(\overline{\ell})$ from $\overline{\ell}$. However, by Proposition 3.4.2, the only minimal attainable sequence for $\mathrm{rot}(\overline{\ell})$ is $(+,-)$, so $\overline{\ell}$ is not rotatable.

Example 3.6.3. Let $\overline{\ell}$ be the annular 1-tangle corresponding to the translation-invariant $\mathbb{Z}$-tangle ℓ shown in Figure 3.5. As shown in the diagram, ℓ has spanning disks realizing $(+,-)$ and $(-,+)$ as attainable sign sequences. The index polynomial is $1-t$, showing that

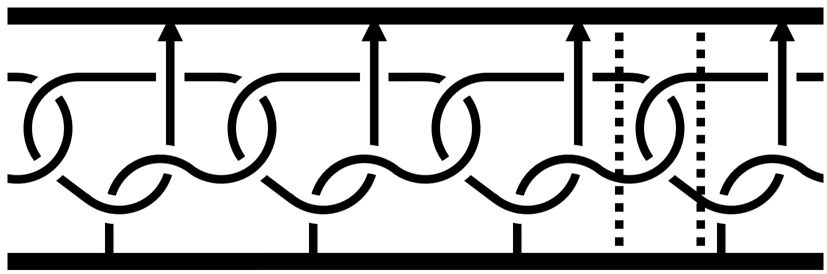

Figure 3.5: A translation-invariant $\mathbb{Z}$-tangle. The corresponding annular 1-tangle has height pair $(1,1)$ with $(+,-)$ and $(-,+)$ both attainable. This annular 1-tangle is reversible.

$\overline{h}_+(\overline{\ell}) = 1$. A lower bound of 1 for $\overline{h}_-(\overline{\ell})$ is provided by the Turaev polynomial: the u^2 coefficient is $-A^{-2} + 2A^{-6} - A^{-10}$. Therefore, $(+,-)$ and $(-,+)$ are both minimal.

The information above gives us an easy way of showing that $\overline{\ell}$ is prime: Since the underpass closure of $\overline{\ell}$ is trivial, we know that $\overline{\ell}$ has no knot-type factor, so to be composite it would have to be a product of two proper annular 1-tangles. One would have to have height pair $(1,0)$, and the other $(0,1)$, but then by Theorem 3.4.5, only one of $(+,-)$ or $(-,+)$ would be attainable for $\overline{\ell}$.

In general, an annular 1-tangle of height 2 falls into one of five categories based on whether its set of minimal attainable sequences is $\{(+,+)\}$, $\{(-,-)\}$, $\{(+,-)\}$, $\{(-,+)\}$, or $\{(+,-),(-,+)\}$. We will further divide the last category into two subcategories.

By Lemma 3.2.3, if $\overline{\ell}$ is an annular 1-tangle of height 2 and both $(+,-)$ and $(-,+)$ are attainable sequences, then there are disjoint spanning disks $\overline{D}_1$ and $\overline{D}_2$ respectively realizing those two sequences as attainable.

Theorem 3.6.4. *Suppose $\overline{\ell}$ is an annular 1-tangle as above. Then exactly one of the following is true.*

(a) The disks $\overline{D}_1$ and $\overline{D}_2$ can be chosen in such a way that both of the intersections of $\overline{\ell}$ with $\overline{D}_1$ come before the intersections with $\overline{D}_2$.

(b) The disks $\overline{D}_1$ and $\overline{D}_2$ can be chosen in such a way that both of the intersections of $\overline{\ell}$ with $\overline{D}_2$ come before the intersections with $\overline{D}_1$.

To prove Theorem 3.6.4, we will use a particular notion of splitting for 2-tangles, analogous to splitting of links: Suppose we have a closed interval $I' \subset \mathbb{R}$ and a 2-tangle in $I' \times I^2$. In this section, the 2-tangle will be called *split* if there is an embedded disk in $I' \times I^2$, disjoint from the strands of the tangle, whose boundary is of the form $\{x\} \times \partial(I^2)$.

Lemma 3.6.5. *Suppose T is a 2-tangle formed from two other tangles T_1 and T_2 in the way shown in Figure 3.6. Then T is split if and only if T_1 is split.*

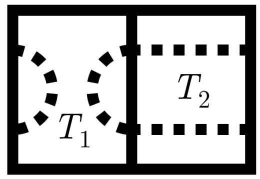

Figure 3.6: A 2-tangle T formed as a sum of two other 2-tangles with particular patterns. The ambient ball $I' \times I^2$ for T is the union $(I'_1 \times I^2) \cup (I'_2 \times I^2)$ of the ambient balls for T_1 and T_2, and the upper bound for I'_1 equals the lower bound for I'_2.

Of course, Lemma 3.6.5 also applies when the the tangles are attached in the other order.

Proof. Of course, if T_1 is split then T is also split. Conversely, suppose that T is split. Then any loop of the form $\{x\} \times \partial(I^2)$ for $x \in I'_1$ is nullhomotopic in the complement of T in $I' \times I^2$. The punctured surface $(I'_1 \cap I'_2) \times I^2) \setminus T$ is incompressible in $(I'_2 \times I^2) \setminus T_2$, because an essential embedded circle on $(I'_1 \cap I'_2) \times I^2) \setminus T$ is clearly not nullhomologous in the larger space. Therefore, the surface is π_1-injective in $(I'_2 \times I^2) \setminus T_2$. Then by Van Kampen's theorem, each $\{x\} \times \partial(I^2)$ is nullhomotopic in $(I'_1 \times I^2) \setminus T_1$. With Dehn's lemma, this implies that there is an embedded disk realizing T_1 as split. $\square$

For an annular 1-tangle $\overline{\ell}$ and spanning disk $\overline{D}$ with $\mathrm{Seq}(\overline{\ell}, \overline{D}) = (+, -)$, performing an ambient isotopy to straighten $\overline{D}$ and then deleting a regular neighborhood of D creates a 3-tangle of the pattern shown in Figure 3.7. This 3-tangle is well-defined as a function of $\overline{\ell}$ and $\overline{D}$ only up to isotopy of $I' \times I^2$ fixing $I' \times \partial(I^2)$, so there can be braiding on the sides. The strands can be labelled as the first, second, and third strands based on the order they appear on $\overline{\ell}$. Let $\mathrm{ST}^3_{\overline{D}}$ be the 2-tangle formed by deleting the first strand, and $\mathrm{ST}^1_{\overline{D}}$ the 2-tangle formed by deleting the third. We will call $\overline{D}$ *(a)-split* if $\mathrm{ST}^3_{\overline{D}}$ is nonsplit and $\mathrm{ST}^1_{\overline{D}}$ is split. Conversely, $\overline{D}$ will be called *(b)-split* if $\mathrm{ST}^3_{\overline{D}}$ is split and $\mathrm{ST}^1_{\overline{D}}$ is nonsplit. For spanning disks $\overline{D}$ with sequence $(-, +)$, we can form $\mathrm{ST}^3_{\overline{D}}$ and $\mathrm{ST}^1_{\overline{D}}$ in a similar way, but we use the opposite convention for (a)- and (b)-splitting: $\overline{D}$ is (a)-split if $\mathrm{ST}^3_{\overline{D}}$ is split and $\mathrm{ST}^1_{\overline{D}}$ is nonsplit.

Note that (a)- and (b)-splitting only apply to pairs $(\overline{\ell}, \overline{D})$ with $\overline{h}_\pm(\overline{\ell}, \overline{D}) = 1$, and no disk may be both (a)-split and (b)-split.

Lemma 3.6.6. *If $\overline{\ell}$ is an annular 1-tangle such that condition (a) from Theorem 3.6.4 is true, then every minimal spanning disk for $\overline{\ell}$ is (a)-split. If instead (b) is true, every minimal spanning disk is (b)-split.*

Proof. Suppose that (a) is true of $\overline{\ell}$. Then $s(\overline{\ell})$ can be drawn as in Figure 3.8, and the tangle T_1 must be nonsplit, as otherwise $\overline{\ell}$ would have height 0. The 3-tangles coming from $\overline{D}_1$

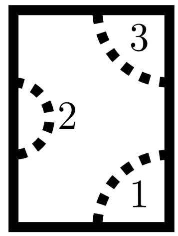

Figure 3.7: A 3-tangle corresponding to a spanning disk with sign sequence $(+, -)$.

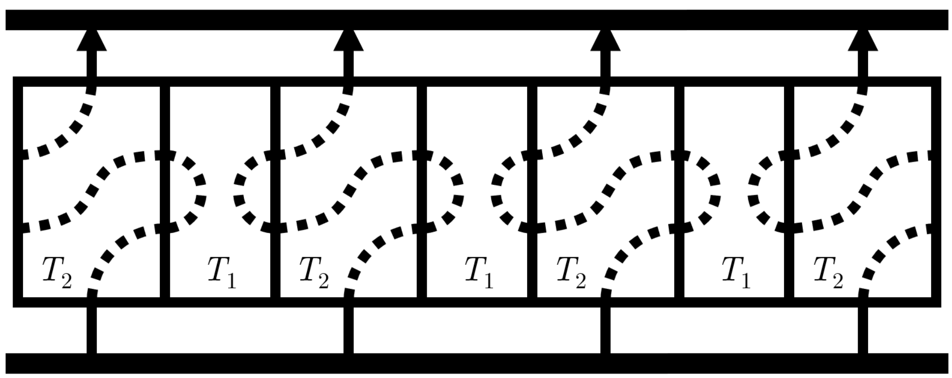

Figure 3.8: An annular 1-tangle for which (a) is true, decomposed into tangles T_1 and T_2 and drawn as a translation-invariant $\mathbb{Z}$-tangle.

and $\overline{D}_2$ are each formed by combining one copy of T_1 with one copy of T_2 in the appropriate way. By Lemma 3.6.5, and the fact that each $\mathrm{ST}^1_{\overline{D}_i}$ is clearly split, $\overline{D}_1$ and $\overline{D}_2$ are both (a)-split. Furthermore, by another application of Lemma 3.6.5, if $\overline{D}$ and $\overline{D}'$ are any two disjoint spanning disks such that one is (a)-split, then the other is (a)-split as well. Then Lemma 3.2.3 implies that all minimal spanning disks are (a)-split.

The same reasoning shows that if (b) is true of $\overline{\ell}$, then all minimal spanning disks of $\overline{\ell}$ are (b)-split. □

Proof of Theorem 3.6.4. First we show that (a) or (b) is true. Suppose we have any choice of $\overline{D}_1$ and $\overline{D}_2$. By the same reasoning as in the proof of Lemma 3.3.2, a positive intersection of $\overline{\ell}$ with $\overline{D}_1$ must be followed by a negative $\overline{D}_1$ intersection or positive $\overline{D}_2$ intersection, and a negative $\overline{D}_2$ intersection must be followed by a positive $\overline{D}_2$ intersection or negative $\overline{D}_1$ intersection. Therefore, the overall sequence of intersections is either $(+\overline{D}_1, -\overline{D}_1, -\overline{D}_2, +\overline{D}_2)$ or $(-\overline{D}_2, +\overline{D}_2, +\overline{D}_1, -\overline{D}_1)$.

That (a) and (b) cannot both be true follows from Lemma 3.6.6. □

Example 3.6.7. Consider the knotoid k shown in Figure 3.9 and its corresponding $\overline{\ell} \in \overline{A}$.

Figure 3.9: A knotoid for which $(-,+)$ is the only minimal attainable sign sequence. It is unlike the Kinoshita knotoid in that its index polynomial is trivial.

The index polynomial is 0, but the Turaev polynomial tells us that the positive and negative heights are both 1. The spanning disk corresponding to the marked shortcut is neither (a)-split nor (b)-split, because the corresponding tangles ST^3_D and ST^1_D are both split. Therefore, $\overline{\ell}$ is neither type (a) nor type (b), so $(-,+)$ is its only minimal attainable sequence.

We now have a partition of the set of height-2 annular 1-tangles into six categories: Type (a), type (b), and four categories for annular 1-tangles that each have only one minimal attainable sequence. For any $\overline{\ell} \in \overline{A}$ with height 2, the rotation $\mathrm{rot}(\overline{\ell})$ is in a different category from $\overline{\ell}$. Together with Corollary 3.4.3, this implies the following corollary.

Corollary 3.6.8. *No proper annular* 1*-tangle with height below* 4 *is rotatable.*

The author does not know if any proper rotatable annular 1-tangles exist. In [7] (Theorem 1.4) it is shown that a knotoid cannot be rotatable if its double branched cover (see Section 2.5) is hyperbolic.

Chapter 4

Braids and Tangles

4.1 Cosmetic Braids on Tangles

In this section, we address two problems about moving the endpoints of a tangle to create a (potentially) new tangle.

Cosmetic Braids on String Links

It is a standard fact about string links that pure braids can be thought of as special string links despite not formally being the same type of object; specifically, pure braids correspond in some way to string links that have inverses ([26], page 392). The same can be said of the relationship between $\mathbb{Z}$-braids and $\mathbb{Z}$-tangles. Since we have been formally treating braids and $\mathbb{Z}$-braids as elements of mapping class groups, we want to establish a convention for treating braids and $\mathbb{Z}$-braids as string links and $\mathbb{Z}$-tangles, so we will do so with a function BT.

In the case of $\mathbb{Z}$-braids and $\mathbb{Z}$-tangles, we define $\mathrm{BT}\colon B/{\sim} \to A/\mathcal{I}$ as follows: Note that any $b \in B$, viewed as a boundary-preserving automorphism of $\mathbb{R}\times I$, is isotopic to the identity automorphism. Given a class in $B/{\sim}$, let us choose a representative $b \in B$ and then choose an isotopy from b to the identity. Let us consider the isotopy to be over times in $[-1, 1]$, where it is b at time -1 and the identity at time 1. Then we can construct a $\mathbb{Z}$-tangle ℓ whereby for each (n, t), we set $\ell(n, t)_y = 2t - 1$ and the x and z coordinates of $\ell(n, t)$ are determined by the position of $(n, 0)$ under the isotopy at time t. The isotopy class of the resulting $\mathbb{Z}$-tangle doesn't depend on our choices, so BT is well-defined. Furthermore, BT is a monoid homomorphism. It is fairly clear that BT commutes with τ, that is, $\mathrm{BT}(\tau[b]) = \tau(\mathrm{BT}([b]))$ for all $[b]$.

Note that not every $\mathbb{Z}$-tangle ℓ with $\ell(n, t)_y = t$ for all n, t is isotopic to a $\mathbb{Z}$-tangle coming from a $\mathbb{Z}$-braid.

For pure braids with finitely many strands, we will still write BT, and the definition carries over without modification. Below, we write $[\ell]$ for the isotopy class of a string link ℓ.

Theorem 4.1.1. *For string links ℓ_1 and ℓ_2, each with k strands, and a nontrivial pure braid of k strands represented by some b, $[\ell_1]\,\mathrm{BT}([b])[\ell_2]$ is not equal to $[\ell_1][\ell_2]$.*

In the special case that ℓ_1 or ℓ_2 is trivial, $[\ell_1]\,\mathrm{BT}([b])$ (for example) can be created from $[\ell_1]$ by composing with a diffeomorphism $\mathbb{R}\times I^2 \to \mathbb{R}\times I^2$ that relabels $\mathbb{R}\times\{1\}\times I$ and fixes the other three faces of the boundary pointwise.

Theorem 4.1.1 can be used to affirmatively answer the question on page 13 of [12]: While Corollary 3.11 in that paper states that the string links K and P are braid-equivalent, the proof in Corollary 3.10 also implies the slightly stronger fact that the braiding in the braid equivalence can be taken to only happen on one side. Therefore, in the notation of [12], we have $K\#L = K\#B\#L$ and $P = K\#B$ for some braid B. At this point we can apply Theorem 4.1.1 to obtain $B = 1$ and hence $K = P$.

We expect that Theorem 4.1.1, together with the results and methods in [12] and [48], can be used to prove the string link monoids are cancellative for all k.

Later we will use this result in the proofs of Theorem 4.3.5 and Theorem 5.6.1 below.

Proof. Suppose we are given a string link ℓ with k strands. For $n > 1$, we will define an isomorphism $\mathrm{Fr}_k/(\mathrm{Fr}_k)_n$ as follows, where Fr_k is the free group on k generators and $(\mathrm{Fr}_k)_n$ is the n^{th} term in the lower central series.

Let M denote the complement of ℓ in $\mathbb{R}\times I^2$, and for $i = \pm 1$ let

$$Q_i = M \cap (\mathbb{R}\times\{i\}\times I) = (\mathbb{R}\times\{i\}\times I)\setminus(\{1,\dots,k\}\times\{i\}\times\{0\}).$$

We can consider each $Q_{\pm 1}$ to have a basepoint at $(0,\pm 1,-1)$ for the purposes of defining fundamental groups. For M, let us use $\{0\}\times I\times\{-1\}$ as an extended basepoint, since the fundamental groups of M with all of these basepoints can be mutually identified by freely travelling along that line segment. Using these basepoints, the inclusion maps $Q_{\pm 1}\to M$ induce maps $\pi_1(Q_{\pm 1})\to\pi_1(M)$.

In [26] (page 393), it is noted that, by a result of Stallings ([46], Theorem 5.1), for any string link ℓ with k strings, for all finite $n > 1$ when we divide by the lower central series the induced maps $\pi_1(Q_{\pm 1})/\pi_1(Q_{\pm 1})_n \to \pi_1(Q_{\pm 1})/\pi_1(Q_{\pm 1})_n$ are isomorphisms. (To apply Theorem 5.1 of [46], the induced maps $H_1(Q_i)\to H_1(M)$ must be isomorphisms and the induced maps $H_2(Q_i)\to H_2(M)$ must be surjective; a standard application of the Mayer–Vietoris theorem shows that $H_1(M)$ is freely generated by meridians of the strands and $H_2(M)$ is trivial. Note that we are using $p = 0$.)

We can cannonically identify Q_{-1} with Q_1, and $\pi_1(Q_{-1})$ and $\pi_1(Q_1)$ with Fr_k. Given ℓ and n, let $\mathrm{Iso}_n(\ell)$ be the composite isomorphism

$$\frac{\mathrm{Fr}_k}{(\mathrm{Fr}_k)_n}\to\frac{\pi_1(Q_1)}{\pi_1(Q_1)_n}\to\frac{\pi_1(M)}{\pi_1(M)_n}\to\frac{\pi_1(Q_{-1})}{\pi_1(Q_{-1})_n}\to\frac{\mathrm{Fr}_k}{(\mathrm{Fr}_k)_n}.$$

Of course, $\mathrm{Iso}_n(\ell)$ will equal $\mathrm{Iso}_n(\iota\circ\ell)$ for all $\iota\in\mathcal{I}$. Thus the maps Iso_n can be considered to form a family of string links invariants. These invariants are all functorial, satisfying $\mathrm{Iso}_n(\ell_1\ell_2) = \mathrm{Iso}_n(\ell_1)\,\mathrm{Iso}_n(\ell_2)$ for all ℓ_1, ℓ_2.

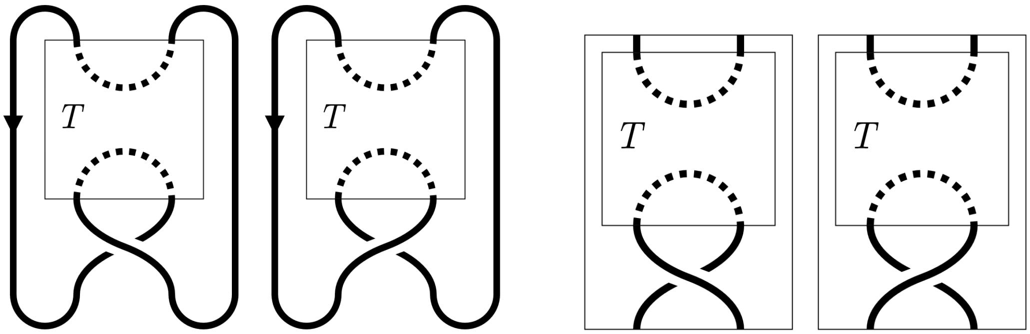

Figure 4.1: On the left, two oriented knots $\mathrm{cl}_+(T)$ and $\mathrm{cl}_-(T)$ constructed as closures of a 2-tangle T. On the right, two new 2-tangles $\mathrm{tw}_+(T)$ and $\mathrm{tw}_-(T)$ constructed from T.

Under the assumptions of our theorem, suppose we have a string link ℓ representing $\mathrm{BT}([b])$. It will suffice to show that there is some n for which $\mathrm{Iso}_n(\ell_1\ell\ell_2)$ is not equal to $\mathrm{Iso}_n(\ell_1\ell_2)$, which, by functoriality, is equivalent to the nontriviality of $\mathrm{Iso}_n(\ell)$.

Since ℓ comes from a braid, the $\mathrm{Iso}_n(\ell)$ are all quotients of the isomorphism on $\pi_1(Q_{\pm 1}) \cong \mathrm{Fr}_k$ induced by b. Since b is nontrivial, the isomorphism on Fr_k is also nontrivial, because if some γ_i acted upon nontrivially by b then the standard generator in Fr_k corresponding to γ_i will be mapped to a nontrivial conjugate of itself. Therefore, there is a $g \in \mathrm{Fr}_k$ with $b_*(g) \neq g$.

A theorem of Magnus ([38], Theorem VIII) holds that free groups are residually nilpotent, and hence every nontrivial element is contained in only finitely many terms of the lower central series (see also [45], Theorem 6.1.10). Therefore there is some n for which $b_*(g)g^{-1}$ represents a nontrivial element of $\mathrm{Fr}_k / (\mathrm{Fr}_k)_n$, which also implies that g and $b_*(g)$ will represent distinct elements. The isomorphism $\mathrm{Iso}_n(\ell)$ maps one of these distinct elements to the other, so it is nontrivial. This completes our proof. □

Cosmetic Crossings on 2-Tangles

A well-known conjecture, the cosmetic crossing conjecture, posits that if a crossing change on an oriented knot can only change the isotopy class of the knot if the switched crossing was *nugatory*. Another way of stating the conjecture is as follows: If a 2-tangle T fitting into the slots in Figure 4.1 (with the two top ends connected and two bottom ends connected) makes the two oriented knots $\mathrm{cl}_\pm(T)$ on the left equivalent, then T must be horizontally split, that is, it admits a disk that that doesn't intersect either strand and has the equator as its boundary. (The two statements are equivalent because everything outside the slot for the tangle is just a neighborhood of the crossing.)

The result below is a weaker alternative version of the cosmetic crossing conjecture:

Theorem 4.1.2. *Suppose that T is a 2-tangle with the top two ends are connected and the bottom two ends connected, and that the two 2-tangles* $\mathrm{tw}_+(T)$ *and* $\mathrm{tw}_-(T)$ *formed from T as in the right side of Figure 4.1 are equivalent. Then T is horizontally split.*

A proof of the standard conjecture would easily extend to this weaker version, because the knots $\mathrm{cl}_+(T)$ and $\mathrm{cl}_-(T)$ can be formed from $\mathrm{tw}_+(T)$ and $\mathrm{tw}_-(T)$, respectively, via the same closure. Therefore, it is easy to see that whenever $\mathrm{tw}_+(T)$ and $\mathrm{tw}_-(T)$ are equivalent to each other, $\mathrm{cl}_+(T)$ and $\mathrm{cl}_-(T)$ are as well.

Proof. We will rely on Theorem 1 from [40].

By the assumption that $\mathrm{tw}_+(T)$ and $\mathrm{tw}_-(T)$ are equivalent, there is a boundary-fixing diffeomorphism $\varphi_1\colon B^3 \to B^3$ that sends $\mathrm{tw}_+(T)$ to $\mathrm{tw}_-(T)$. There is also a simple diffeomorphism $\varphi_2\colon B^3 \to B^3$ that, like φ_1, sends $\mathrm{tw}_+(T)$ to $\mathrm{tw}_-(T)$ but whose restriction to ∂B^3 acts as a Dehn twist around the equator. On the interior of B_3, φ_2 acts as a Dehn twist around a disk in the sense of [40] (page 1333).

Now let us consider $\varphi_2^{-1} \circ \varphi_1\colon B^3 \to B^3$. This composition sends $\mathrm{tw}_+(T)$ to itself and the restriction to ∂B^3 is again a Dehn twist around the equator. In ∂B^3, the diffeomorphism leaves a neighborhood of the four endpoints of T fixed pointwise. To $\varphi_2^{-1} \circ \varphi_1$ we can associate a pair of integers, one for each component of T, expressing the "framing change", as follows: Consider the vector bundle whose base space is T (that is, the space of points on the two strands) and whose fiber at each point is the quotient of the 3-dimensional tangent space in B^3 by the 1-dimensional tangent space in T. The derivative of $\varphi_2^{-1} \circ \varphi_1$ induces an automorphism of this vector bundle that restricts to the identity on the endpoints. Therefore there is some total integral amount of twisting on each strand. Note that, for each individual strand, it is possible to isotope φ_2 (allowing the boundary to move) to the identity without moving a neighborhood of the endpoints of that chosen strand. Therefore, for each strand, $\varphi_2^{-1} \circ \varphi_1$ will have the same framing change as a boundary-fixing isotopy from a 1-tangle to itself, which is 0.

The fact that the framing changes are both 0 implies that $\varphi_2^{-1} \circ \varphi_1$ can be perturbed in a neighborhood of T, without moving ∂B^3 to an automorphism $\varphi\colon B^3 \to B^3$ that fixes a regular neighborhood of the two strands pointwise. Let M be the complement in B^3 of the regular neighborhood chosen above. The restriction of φ to M is a diffeomorphism whose further restriction to ∂M is a Dehn twist. Now we can apply Theorem 1 of [40] to φ and M. The theorem guarantees the existence of a disk in M whose boundary is the equator, which therefore renders $\mathrm{tw}_\pm(T)$ horizontally split. The disk can then pushed inward toward T, so T must be horizontally split as well. □

4.2 Symmetric Braids

This section is on both $\mathbb{Z}$-braids and $\mathbb{N}$-braids, in particular the question of which such braids are equivalent up to isotopy to their translates.

Symmetric $\mathbb{Z}$-Braids

Recall from Section 2.3 that B is the set of automorphisms of $\mathbb{R} \times I$ that fix both $\partial(\mathbb{R} \times I)$ and $\mathbb{Z} \times \{0\}$ pointwise. Note that every (strictly) translation-invariant element of B is in the same class as the identity: As we showed in Section 2.3, $\overline{B}$ is path-connected, and therefore every element of $B_{\mathbb{Z}}$ is isotopic to the identity. We can posit something stronger by showing that $(B/{\sim})_{\mathbb{Z}}$ is trivial.

Theorem 4.2.1. *There is no nontrivial element of $B/{\sim}$ that is fixed by translation.*

Proof. Suppose that we have a $b \in B$ with $b \sim \tau b$. Our goal is to show $b \sim \mathrm{id}$. In order to invoke hyperbolic geometry, we will treat b as a pure automorphism of $(\mathbb{R}\times(-1,1))\setminus(\mathbb{Z}\times\{0\})$, which covers the thrice-punctured sphere $(S^1 \times (-1,1)) \setminus \{(0,0)\}$. In this covering, the upstairs surface fails to be a hyperbolic surface in the strict sense ([19], page 21) because it does not have finite area, but the thrice-punctured sphere is hyperbolic.

Recall that b determines paths $b(\gamma_n)$, where $\gamma_n = \{n\} \times [-1,0]$. Abusing notation, we will also use $b(\gamma_n)$ to refer merely to the part of $b(\gamma_n)$ in $(\mathbb{R} \times (-1,1)) \setminus (\mathbb{Z} \times \{0\})$. The image of each such path in $(S^1 \times (-1,1)) \setminus \{(0,0)\}$ is an immersed curve $\overline{b\gamma_n}$. By the assumption of $b \sim \tau b$, all of the $\overline{b\gamma_n}$ are homotopic (through end-proper paths). Let $\overline{\beta}$ be the geodesic representative of this homotopy class.

We will show now that $\overline{\beta}$ is embedded. Suppose otherwise. Consider the family of lifts β_n to $(\mathbb{R} \times (-1,1)) \setminus (\mathbb{Z} \times \{0\})$. Each β_n is homotopic to $b(\gamma_n)$ and is a translate of β_0. Since $\overline{\beta}$ has a self-intersection, there is a pair of (not necessarily distinct) lifts that intersect each other (or self-intersect in the case of identical lifts). We can take these two lifts to be β_0 and β_m for some m. In any intermediate (thereby finite) cover of the thrice-punctured sphere, the images of β_0 and β_m will still intersect. However, since $b(\gamma_0)$ and $b(\gamma_m)$ are embedded and disjoint and bounded in the second dimension, their images in the intermediate k-fold cover are also embedded and disjoint for large values of k. This contradicts Proposition 1.6 or Corollary 1.9 of [19]. (Those are stated for closed curves but hold equally well for arcs.) Therefore, $\overline{\beta}$ is embedded.

Because $\overline{\beta}$ is embedded, it is the unique embedded geodesic between its endpoints in the thrice-punctured sphere. Thus each $b(\gamma_n)$ is homotopic to γ_n. Therefore, by Proposition 2.3.3, b is isotopic to the identity. □

In contrast, there are $\mathbb{Z}$-tangles ℓ that look like braids in the sense that they satisfy $\ell(n,t)_y = t$ for all n and t, and for which $p(\ell) \in (A/\mathcal{I})_{\mathbb{Z}}$, but which are not isotopic to the trivial $\mathbb{Z}$-tangle. For example, the $\mathbb{Z}$-tangle in Section 6.1 is isotopic to a $\mathbb{Z}$-tangle satisfying the equation above, but in Section 6.1 we also show that that isotopy class of $\mathbb{Z}$-tangles is nontrivial.

Symmetric $\mathbb{N}$-Braids

The following result is the closest statement to Theorem 4.2.1 that can be made for $\mathbb{N}$-braids.

Theorem 4.2.2. *There is no nontrivial isotopy class of pure* $\mathbb{N}$*-braids that is fixed by* τ^{-1}.

We will use a proof that applies the Birman–Hilden map and relies on standard results about knotoids.

Proof. Suppose b is a pure $\mathbb{N}$-braid with $b \sim \tau^{-1}b$. Up to isotopy through $\mathbb{N}$-braids, we can assume without loss of generality that b restricts to the identity on the set $\{x \leq 0\}$.

Both b and $\tau^{-1}b$ fix $\{x \leq -1\}$ pointwise, and nothing of interest happens in $\{x \leq -1\}$, so there exist isotopies realizing $b \sim \tau^{-1}b$ that fix $\{x \leq -1\}$ throughout the time interval. Let us fix such an isotopy $\varepsilon_0 \colon [0,1] \to \{\mathbb{N}\text{-braids}\}$ from $\tau^{-1}b$ to b.

Note that, by our assumption that b fixes $\{x \leq 0\}$ pointwise, b is not just a pure $\mathbb{N}$-braid but also a pure $\mathbb{Z}$-braid. Then for $n \in \{0\} \cup \mathbb{N}$, let us define a new $\mathbb{N}$-braid b_n by $b_n = \tau^n b$. Thus $(\tau^{-1})^n b_n = b$ for all n.

The definition of BT extends naturally to $\mathbb{N}$-braids: As in the original definition, there are isotopies from b to the identity that, while not necessarily fixing the marked points, fix $\partial(\mathbb{R} \times I)$ throughout the time interval, and each such isotopy can be used to interpret b as a tangle (in this case, as a $\mathbb{N}$-*tangle*).

In order to apply BT, we will choose a specific isotopy

$$\varepsilon \colon [-1,1] \to \{\partial\text{-fixing automorphisms of } \mathbb{R} \times I\}$$

from b to the identity by adapting the "combed" form for standard pure braid groups ([1], page 395). First we will define the restriction of ε to the time interval $[-1,1)$. Then we will check that our isotopy on $[-1,1)$ limits to the identity and thus extends as desired to I.

For each $n \in \{0\} \cup \mathbb{N}$, let us set $\varepsilon(1 - 2^{1-n}) = b_n$. Now, for each $n \in \mathbb{N}$, let ε_n be the same as e_0 but reparameterized to occur in the time interval $[1 - 2^{2-n}, 1 - 2^{1-n}] \subset [-1,1]$ and shifted via conjugation to the right by n (thus each ε_n is a boundary-fixing isotopy from b_{n-1} to b_n). Then we can legally define ε's behavior on all of $[-1,1)$ by stipulating that its restriction to each $[1 - 2^{2-n}, 1 - 2^{1-n}]$ equals ε_n.

It remains to show that setting $\varepsilon(1)$ to be the identity will make ε continuous on I. According to the compact-open topology, for any compact subset of $\mathbb{R} \times I$ the restrictions of ε to the subset must uniformly converge to the inclusion map into $\mathbb{R} \times I$. Note that each ε_n keeps $\{x \leq n-1\}$ fixed. For any compact region of $\mathbb{R} \times I$, there is a maximal value of x. If n is a natural number for which $n-1$ is greater than that maximal value, then ε will act as the identity on the compact region after time $1 - 2^{2-n}$. Therefore, the limit of ε as time approaches 1 exists and is equal to the identity. This finishes our definition of ε.

Now, abusing notation, let $\mathrm{BT}(b)$ denote the specific $\mathbb{N}$-tangle formed from b via the isotopy ε.

Since b is an automorphism of $\mathbb{R} \times I$, it has an inverse b^{-1}. Let $\mathrm{BT}(b^{-1})$ denote the result of flipping $\mathrm{BT}(b)$ in the y (and t) direction. As the notation suggests, $\mathrm{BT}(b^{-1})$ is a $\mathbb{N}$-tangle realization for b^{-1}. Then the product $\mathrm{BT}(b^{-1})\,\mathrm{BT}(b)$ is a $\mathbb{N}$-tangle isotopic to the trivial $\mathbb{N}$-tangle. (A specific element of $\mathcal{I}$ realizing this relationship can be read off from the previously chosen boundary-fixing isotopy of $\mathbb{R} \times I$ between b and the identity. Specifically, we can use the y values as proxies for the time values of the isotopy by identifying each of $[0,1]$ and $[-1,0]$ with $[-1,1]$ appropriately.) Therefore it admits a family of spanning disks D_n corresponding to the spanning disks $\{n+\frac{1}{2}\} \times I^2$ for the trivial $\mathbb{N}$-tangle. (To illustrate, if we use the aforementioned element of $\mathcal{I}$ to determine the family, then the intersection of each D_n with $\{y=0\}$ will look like $b(\{n+\frac{1}{2}\} \times I)$.) In particular, the spanning disks all have height 0, and the union of the spanning disks is biproperly embedded in $\mathbb{R} \times I^2$.

Note that our $\mathrm{BT}(b)$ is fully determined up to isotopy as a $\mathbb{N}$-tangle just by the path $b(\gamma_1)$. The resulting $\mathbb{N}$-tangle is essentially constructed by a stacked sequence of blocks that are mutually identical up to shifting in the x direction and shifting and scaling in the y direction. (See Figure 4.2, upper half.) Outside of these blocks, $\mathrm{BT}(b)$ behaves in the same way as the trivial $\mathbb{N}$-tangle. With an ambient isotopy that skews all of the blocks to put them all at the same level as each other in the y direction, we obtain a new $\mathbb{N}$-tangle whose strands are all $\mathbb{Z}$-translates of each other. Therefore this new $\mathbb{N}$-tangle can be extended to a translation-invariant $\mathbb{Z}$-tangle $\ell \in A_{\mathbb{Z}}$ as in the lower half of Figure 4.2. Let ℓ^{-1} denote the mirror image of ℓ, obtainable also by skewing and extending $\mathrm{BT}(b^{-1})$. (Note that any $\mathbb{Z}$-tangle can by flipped in the y direction, but only in the case of $\mathbb{Z}$-tangles coming about by BT will the product be isotopic to the identity.)

Consider how the spanning disks (D_n) behave under the isotopy from $\mathrm{BT}(b^{-1})\,\mathrm{BT}(b)$ to the $\mathbb{N}$-tangle obtained from $\ell^{-1}\ell$ by forgetting the nonpositively indexed strands. By properness, there is some n for which the isotoped copy of D_n sits entirely to the right of the rightmost point of the 0^{th} strand of $\ell^{-1}\ell$. Therefore, there is a spanning disk of height 0 for $\ell^{-1}\ell$. By Lemma 3.1.3, this implies that $s^{-1}(\ell^{-1}\ell) = s^{-1}(\ell^{-1})s^{-1}(\ell)$ has height 0. Next, Corollary 3.4.7 implies that $s^{-1}(\ell)$ has height 0, so it is knot-type (see Section 2.5). The oriented knot in question will be equal to $p(\ell)/\infty$ (see Section 2.5), which can be seen to be trivial from the fact that the y coordinates of the strands of $\mathrm{BT}(b)$ are monotone in t.

Now we know that $s^{-1}(\ell)$ is isotopic to the trivial annular 1-tangle, so ℓ is isotopic to the trivial annular 1-tangle. Since the isotopy class of $\mathrm{BT}(b)$ is obtainable by forgetting strands of ℓ, it follows that $\mathrm{BT}(b)$ is the trivial $\mathbb{N}$-tangle. This implies that the first strand does not braid nontrivially around the other strands, since each strand can be read as representing an element of the fundamental group of the complement of the other strands, and that complement retracts onto the part of the complement in $\mathbb{R} \times \{-1\} \times I$. Therefore, $b(\gamma_1)$ is isotopic to γ_1, and this determines the rest of b up to isotopy, as we argued earlier. Hence, we are done. □

4.3 Decomposing Ribbons

This section builds toward Theorem 4.3.5, which addresses the geometric aspects of the stacking operation for $(A/\mathcal{I})_{\mathbb{Z}}$ as a monoid. The result does not generalize to $A/\mathcal{I}$.

First, we observe that $\mathbb{Z}$-braids are determined by their finite sub-braids:

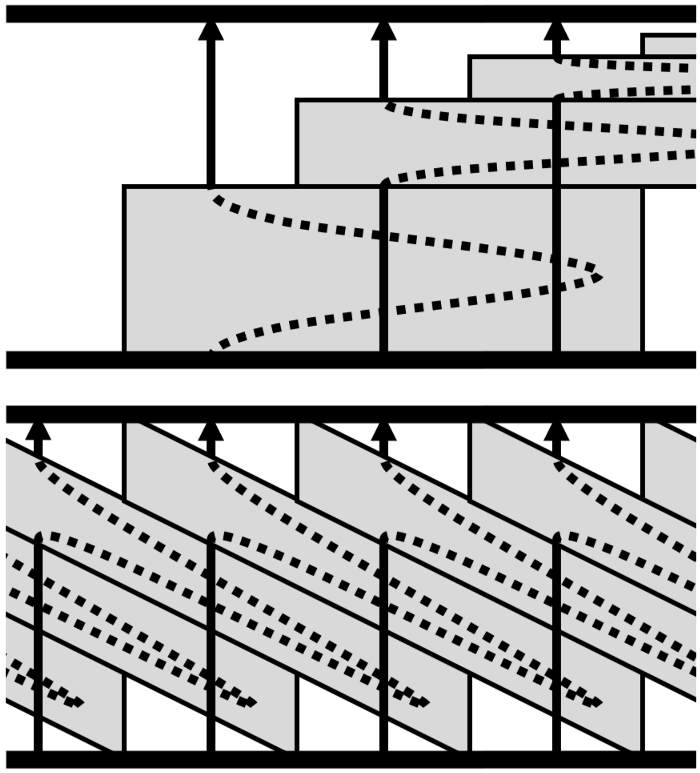

Figure 4.2: Creating a symmetric $\mathbb{Z}$-tangle $\ell \in A_{\mathbb{Z}}$ from a $\mathbb{N}$-braid that is symmetric up to isotopy.

Lemma 4.3.1. *Suppose $[b] \in B/\!\sim$ is a nontrivial $\mathbb{Z}$-braid. Then there is a finite subset of $\mathbb{Z}$ such that the finite-stranded braid obtained by forgetting all other strands of $[b]$ is also nontrivial.*

Proof. For such a b, by Proposition 2.3.3, there is a n such that $b(\gamma_n)$ is not homotopic to γ_n. Fix a $k \in \mathbb{N}$ such that k is greater than the maximum absolute value of all x coordinates of points on $b(\gamma_n)$.

We claim that there is no based homotopy from $b(\gamma_n)$ to γ_n even if the homotopy is allowed to pass through punctures besides those indexed $-k, \ldots, k$. Supposing such a homotopy existed, we could choose an identification of the open intervals (k, ∞) and $(k, k+1)$ and apply it to the homotopy to get a new homotopy that occurs entirely within the space with $-k-1 < x < k+1$.

Therefore the corresponding braid with $2k+1$ strands is nontrivial. □

We can use Lemma 4.3.1 to extend Theorem 4.1.1 to the context of $\mathbb{Z}$-braids:

Corollary 4.3.2. *For $\mathbb{Z}$-tangles ℓ_1 and ℓ_2 and a nontrivial class of $\mathbb{Z}$-braids $[b] \in B/\!\sim$, $p(\ell_1)\,\mathrm{BT}([b])p(\ell_2)$ is not equal to $p(\ell_1)p(\ell_2)$.*

Proof. Under the stated assumptions, Lemma 4.3.1 implies that there is a finite subset of $\mathbb{Z}$ such that the braid obtained from $[b]$ by forgetting all other strands is nontrivial. Then by Theorem 4.1.1, the string links coming from $p(\ell_1)\,\mathrm{BT}([b])p(\ell_2)$ and $p(\ell_1)p(\ell_2)$ using that same subset are distinct. Therefore $p(\ell_1)\,\mathrm{BT}([b])p(\ell_2)$ and $p(\ell_1)p(\ell_2)$ are not equal. □

Since BT is a monoid homomorphism, the behavior of the expression when either $p(\ell_i)$ is set to be trivial can be described as a group action. That is, we can define left- and right actions, respectively, of $B/\sim$ on $A/\mathcal{I}$ by $([b], p(\ell)) \mapsto \mathrm{BT}([b])p(\ell)$ and $(p(\ell), [b]) \mapsto p(\ell)\,\mathrm{BT}([b])$.

Corollary 4.3.3. *The two braiding actions above are free.*

Proof. This follows immediately from Corollary 4.3.2. □

Lemma 4.3.4. *For all $p(\ell) \in (A/\mathcal{I})_{\mathbb{Z}}$ and $[b] \in B/\sim$, if $\mathrm{BT}([b])p(\ell)$ or $p(\ell)\,\mathrm{BT}([b])$ is in $(A/\mathcal{I})_Z$ then $[b]$ is the identity.*

Proof. Suppose we have ℓ and b with $p(\ell) = \tau p(\ell)$ and $\mathrm{BT}([b])p(\ell) = \tau(\mathrm{BT}([b])p(\ell))$. Then we have the following equalities:

$$\mathrm{BT}([b])p(\ell) = \tau\,\mathrm{BT}([b])p(\ell) = (\tau\,\mathrm{BT}([b]))(\tau p(\ell)) = \mathrm{BT}(\tau[b])p(\ell)$$

Above, the first equality is by assumption, the second expresses the fact that τ is a monoid homomorphism, and the third uses the fact that τ commutes with BT and the assumption that $p(\ell) = \tau p(\ell)$.

By Corollary 4.3.3, it follows that $[b] = \tau[b]$ and therefore that $[b]$ is the identity, by Theorem 4.2.1.

The same reasoning works in the case that $p(\ell)\,\mathrm{BT}([b]) = \tau(p(\ell)\,\mathrm{BT}([b]))$. □

Theorem 4.3.5. *Suppose Q is a decomposing ribbon for some $\mathbb{Z}$-tangle $\ell \in A$ with $p(\ell) \in (A/\mathcal{I})_{\mathbb{Z}}$. Suppose furthermore that $\ell_1, \ell_2, \ell_1', \ell_2' \in A$ and $\iota, \iota' \in \mathcal{I}$ satisfy $\ell = \iota \circ (\ell_1\ell_2) = \iota' \circ (\ell_1'\ell_2')$ and $Q = \iota(\mathbb{R} \times \{0\} \times I) = \iota'(\mathbb{R} \times \{0\} \times I)$. If $p(\ell_1), p(\ell_2), p(\ell_1'), p(\ell_2') \in (A/\mathcal{I})_{\mathbb{Z}}$, then $p(\ell_1) = p(\ell_1')$ and $p(\ell_2) = p(\ell_2')$.*

Theorem 4.3.5 says that in the context of $\mathbb{Z}$-tangles that are translation-invariant up to isotopy, a valid decomposing ribbon (that is, one for which each side can be interpreted as representing a translation-invariant class) expresses a $\mathbb{Z}$-tangle as a specific product. Note that the product of elements of $A/\mathcal{I}$ not in $(A/\mathcal{I})_{\mathbb{Z}}$ can still itself be in $(A/\mathcal{I})_{\mathbb{Z}}$.

Proof. Under the hypotheses of the theorem, there is a diffeomorphism of $\mathbb{R} \times I^2$ that sends ℓ_1 to ℓ_1' and is the identity on $(\mathbb{R} \times I \times \{\pm 1\}) \cup (\mathbb{R} \times \{-1\} \times I)$. The restriction of the homeomorphism to $\mathbb{R} \times \{1\} \times I$ defines a $\mathbb{Z}$-braid, which, when acting on $p(\ell_1)$ on the right, returns $p(\ell_1')$. Then by Lemma 4.3.4, the braid is trivial. Therefore, ℓ_1 and ℓ_1' are isotopic. For the same reasons but with left action, we can say also that ℓ_2 is isotopic to ℓ_2'. □

Chapter 5

Birman–Hilden Theory

In this chapter we focus on the Birman–Hilden map Bh, in particular the question of whether it is injective. An affirmative answer would signify that knotoids could be thought of as a subtype of $\mathbb{Z}$-tangles up to isotopy. With this question in mind, we will give a number of sufficient conditions on pairs of distinct elements $q(\ell_1), q(\ell_2)$ of $A_{\mathbb{Z}}/\mathcal{I}_{\mathbb{Z}}$ that ensure that they have distinct images under Bh.

As we saw in Lemma 3.1.3, height is an example of an invariant that is remembered by Bh, since $h \circ \mathrm{Bh} = \overline{h} \circ s_*^{-1}$. The same is true of the signed heights. Thus, for two annular 1-tangles to have the same image under the Birman–Hilden map they must have the same signed heights.

If A is given a topology whereby isotopies are the same as paths in A, then $A/\mathcal{I}$ is the set of path components and similarly $A_{\mathbb{Z}}/\mathcal{I}_{\mathbb{Z}}$ is the set of path components of $A_{\mathbb{Z}}$. Thus, to say that the map $\mathrm{Bh}\colon A_{\mathbb{Z}}/\mathcal{I}_{\mathbb{Z}} \to p(A_{\mathbb{Z}})$ is injective is equivalent to saying that the inclusion $A_{\mathbb{Z}} \to A$ induces an injection on π_0. One might hope that the inclusion of any component of $A_{\mathbb{Z}}$ into the corresponding component of A is, say, a homotopy equivalence; however, this will not be true in general. For example, if ℓ is a nontrivial knot-type $\mathbb{Z}$-tangle and ι is an isotopy from ℓ to itself resulting from locally performing a nontrivial symmetry on a single strand, then ι is not connected to any element of $\mathcal{I}_{\mathbb{Z}}$ through any 1-dimensional family of isotopies fixing ℓ.

We will address the question of surjectivity of the Birman–Hilden map (that is, whether $p(A_{\mathbb{Z}})$ is a proper subset of $(A/\mathcal{I})_{\mathbb{Z}}$) in Section 6.1.

5.1 The Affine Index Polynomial

Recall that the affine index polynomial from Section 3.5 and [23] is an easy-to-compute invariant of annular 1-tangles. In this section we will show that the affine index polynomial of an annular 1-tangle $\overline{\ell} \in \overline{A}$ is determined by $p(s(\overline{\ell})) = \mathrm{Bh}(s_*(\overline{q}(\ell)))$.

For a $\mathbb{Z}$-tangle $\ell \in A$, any finite set of strands can be turned into an oriented link: Considering $\mathbb{R} \times I^2$ as a subset of $\mathbb{R}^3$, for each $x \in \mathbb{Z}$ with a corresponding strand under consideration let us close the strand with an arc from $(x, 1, 0)$ to $(x, -1, 0)$ in $\{x\} \times (R^2 \setminus I^2)$. (It does not matter whether the new arc goes over or under $\mathbb{R} \times I^2$.) The resulting link is, of course, an isotopy invariant of ℓ. Given $\overline{\ell} \in \overline{A}$ and any $k \in \mathbb{N}$, it is straightforward to check from the definition of the affine index polynomial that the $t^{\pm k}$ coefficient of AIP for the knotoid corresponding to $\overline{q}(\overline{\ell})$ equals the linking number of the link coming from two components of $s(\overline{\ell})$. (Alternatively, we could use two adjacent components of $s(\overline{\ell})/k$.) Therefore, the affine index polynomial factors through the Birman–Hilden map.

In contrast, it is not obvious one way or the other that the same property holds of the index polynomial from [30].

5.2 Diffeomorphisms Moving the Boundary

Recall from Section 2.5 that there is a superset $\mathcal{J}$ of $\mathcal{I}$ acting on A, with a subset $\mathcal{J}_{\mathbb{Z}}$ acting on $A_{\mathbb{Z}}$.

Proposition 5.2.1. *If $\ell_1, \ell_2 \in A_{\mathbb{Z}}$ are translation-invariant $\mathbb{Z}$-tangles in the same orbit under $\mathcal{J}$ and if $q(\ell_1)$ is uniquely determined by its image under* Bh*, then so is $q(\ell_2)$.*

Proof. By assumption, there is a $\eta \in \mathcal{J}_{\mathbb{Z}}$ such that $\eta \circ \ell_1 = \ell_2$, and for all $\iota_1 \in \mathcal{I}$ and $\ell_3 \in A_{\mathbb{Z}}$ with $\iota_1 \circ \ell_1 = \ell_3$, there exists a $\iota_1' \in \mathcal{I}_{\mathbb{Z}}$ with $\iota_1' \circ \ell_1 = \ell_3$.

Suppose, for some $\iota_2' \in \mathcal{I}$ and $\ell_4 \in A_{\mathbb{Z}}$, that $\iota_2 \circ \ell_2 = \ell_4$. Then we have $\iota_2 \circ \eta \circ \ell_1 = \ell_4$ and hence $\eta^{-1} \circ \iota_2 \circ \eta \circ \ell_1 = \eta^{-1} \circ \ell_4$. Setting $\iota_1 = \eta^{-1} \circ \iota_2 \circ \eta$ and $\ell_3 = \eta^{-1} \circ \ell_4$, we see that there is a $\iota_1' \in \mathcal{I}_{\mathbb{Z}}$ with $\iota_1' \circ \ell_1 = \eta^{-1} \circ \ell_4$. Then $\eta \circ \iota_1' \circ \ell_1 = \ell_4$, and hence $\eta \circ \iota_1' \circ \eta^{-1} \circ \ell_2 = \ell_4$, so $q(\ell_2) = q(\ell_4)$. □

The utility of Proposition 5.2.1 comes from the fact that the orbits involved are very large. For instance, a cursory check shows that, of the 31 knotoids (taken up to basic involutions) listed in Figure 1 of [32], over half correspond to elements of $A_{\mathbb{Z}}$ that are all in the same orbit of $\mathcal{J}$, which is also the orbit containing the trivial $\mathbb{Z}$-tangle.

5.3 The Knot-Type Case

Using Lemma 3.1.3, we can make a quick conclusion about the Birman–Hilden invariant behaves in the knot-type case.

Theorem 5.3.1. *If $\ell \in A_{\mathbb{Z}}$ has height* 0*, then $q(\ell)$ is uniquely determined by its image under* Bh*.*

Proof. Given ℓ as above, suppose that some other $\ell' \in A_{\mathbb{Z}}$ has $p(\ell') = p(\ell)$. Then ℓ' also has height 0. By Lemma 3.1.3, $s^{-1}(\ell)$ and $s^{-1}(\ell')$ also have height 0. Therefore, $\overline{q}(s^{-1}(\ell))$ and $\overline{q}(s^{-1}(\ell'))$ correspond to knot-type knotoids $\kappa^\bullet$ and $(\kappa')^\bullet$. The knots κ and κ' are recoverable from $p(\ell)/\infty$ and $p(\ell')/\infty$. Therefore, $\kappa = \kappa'$, so we have $q(\ell) = s_*(\kappa^\bullet) = s_*((\kappa')^\bullet) = q(\ell')$. □

5.4 Products with Knot-Types

We would like to be able to say generally that whenever two translation-invariant $\mathbb{Z}$-tangles both are not counterexamples to injectivity of Bh, their product is not a counterexample either. In this section, we will prove this statement under the extra condition that one of the $\mathbb{Z}$-tangles is knot-type. Later in this chapter, the result will be useful for dealing with the set of annular 1-tangles of height 2.

Lemma 5.4.1. *Given $\ell \in A_{\mathbb{Z}}$, if there exists a ball Ω in $\mathbb{R} \times I^2$ for which $\ell \cap \Omega$ has one component and the oriented knot κ determined by $\ell \cap \Omega \subset \Omega$ is prime, then $\kappa^\bullet$ is a factor of $\overline{q}(s^{-1}(\ell))$.*

The lemma above remains true if we remove the word "prime". For simplicity, we will only prove the prime case, since that is the only case we will need later.

Proof. Our proof strategy is adapted from the proof of Theorem 2.10 in [35].

Choose a minimal spanning disk Δ_0 for ℓ, which by Lemma 3.1.3 can also be chosen to be disjoint from its $\mathbb{Z}$-translates $\Delta_i, i \in \mathbb{Z}$. Let $\Delta \subset \mathbb{R} \times I^2$ denote the union of all of the Δ_i. Since the knot inside Ω is the only property of Ω that matters, we can choose Ω such that (a) $\partial\Omega$ is transverse to Δ, (b) $\partial\Omega \cap \Delta \cap \ell$ is empty, and (c) the number of components of $\partial\Omega \cap \Delta$ is minimal over all choices of Ω.

Consider a component circle of $\partial\Omega \cap \Delta$. Within $\partial\Omega$, the circle either bounds two disks that each intersect ℓ once or bounds a disk disjoint from ℓ and another that intersects ℓ twice. Therefore, the disk in Δ bounded by the same circle intersects ℓ at most once, because otherwise we could swap the disks and apply Lemma 3.1.2 to obtain a contradiction to the minimality of the components of Δ.

Suppose that there is at least one component of $\partial\Omega \cap \Delta$, and suppose that at least one bounds a disk on Δ (or equivalently on $\partial\Omega$) disjoint from ℓ. Then let C be an innermost (on Δ) such circle, which must also be innermost among all components of $\partial\Omega \cap \Delta$. Then the disks on $\partial\Omega$ and Δ bounded by C and disjoint from ℓ form a sphere that in turn bounds an empty ball, contradicting the assumption that Ω was chosen to minimize the number of intersection circles with Δ. Therefore any component of $\partial\Omega \cap \Delta$ loops once around ℓ in Ω.

By the supposition that $\partial\Omega \cap \Delta$ is nonempty, $\Omega \cap \Delta$ is nonempty as well. Suppose that some component is a disk. Then that disk must intersect ℓ in exactly one point. Therefore the disk expresses κ as a connected sum of two knots, one of which must be trivial by our assumption of primality. We are then able to push Ω through the side of the disk with the trivial knot, thus reducing the size of the intersection of Ω with Δ. This contradicts our minimality assumption, so no component of $\Omega \cap \Delta$ is a disk.

Now consider a component of $\Omega \cap \Delta$, which has at least two 1-dimensional boundary components. Each circular component has an unsigned linking number of 1 with ℓ. Therefore, the disks in Δ bounded by all but the "outer" component on Δ intersect ℓ once. The outer circle bounds a disk in $\partial\Omega$ that intersects ℓ once, and each component of Δ is supposed to realize the height of ℓ, so by Lemma 3.1.2, the disk in Δ enclosed by the outer circle only intersects ℓ once. This disk includes the intersections associated to the other circles, so there can only be one other circle. Therefore each component of $\Omega \cap \Delta$ is simply an annulus that doesn't intersect ℓ anywhere.

Consider any such annulus E. By a standard application of Dehn's lemma, the side of E in Ω on which the strand of ℓ lies is a thickened knot in Ω. The knot is a factor of κ, so E either follows ℓ or contains κ inside, following $\partial\Omega$. In either case we can reduce the amount of intersection between Ω and Δ by making the appropriate modification as in [35] in the proof of Theorem 2.10, obtaining a contradiction.

We have deduced that Ω must be disjoint from Δ. Therefore, it is contained in just one component of the complement of Δ in $\mathbb{R} \times I^2$, and that one component is a fundamental domain for translation by integers. Therefore there is a corresponding ball $\overline{\Omega}$ in $S^1 \times I^2$ which encloses κ in a strand of $s^{-1}(\ell)$.

Now it is easy to conclude that $\kappa^\bullet$ is a factor of $\overline{q}(s^{-1}(\ell))$: Let $\overline{\Omega}' \subset S^1 \times I^2$ be the union of $\overline{\Omega}$ with a neighborhood of $S^1 \times \{-1\} \times I$ and with a neighborhood of the part of $s^{-1}(\ell)$ that goes from $(0, -1, 0)$ to $\overline{\Omega}$. The boundary $\partial\overline{\Omega}'$ is a decomposing annulus $\overline{Q}$ expressing $\overline{q}(s^{-1}(\ell))$ as a product with $\kappa^\bullet$ as one factor. □

We will say that a $\mathbb{Z}$-tangle ℓ is *knotless* if the intersection of a ball $\Omega \subset \mathbb{R} \times I^2$ with ℓ is trivial as a tangle in Ω whenever the intersection has exactly one component.

Lemma 5.4.2. *Given $\ell_0, \ell_1, \ell_2 \in A_{\mathbb{Z}}$ such that $\overline{q}(s^{-1}(\ell_0)) = \kappa^\bullet$ for some prime knot κ and $p(\ell_0\ell_1) = p(\ell_0\ell_2)$, it is also true that $p(\ell_1) = p(\ell_2)$.*

Proof. First let us pick an isotopy ι from $\ell_0\ell_2$ to $\ell_0\ell_1$.

Since ℓ_0 is knot-type, there is a ball Ω^1 in $S^1 \times I^2$ for which the intersection with $s^{-1}(\ell_0\ell_1)$ is κ in the form of a 1-tangle in the ball. Therefore, there is a family of pairwise disjoint balls Ω^1_n for $n \in \mathbb{Z}$ in $\mathbb{R} \times I^2$ isolating copies of κ in $\ell_0\ell_1$. These balls can be chosen so that each Ω^1_n intersects only the n^{th} strand of $\ell_0\ell_1$ and such that it is the translate of Ω^1_0 by n. A similar family of balls exists for $\ell_0\ell_2$. For each n, let Ω^2_n denote the result of applying ι to the n^{th} ball in that second family. Therefore the Ω^2_n are like the Ω^1_n in that they are a end-properly embedded family of pairwise disjoint balls with each Ω^2_n intersecting $\ell_0\ell_1$ as the knot κ, but Ω^2_n won't in general be the n^{th} translate of Ω^2_0. We can recover $p(\ell_1)$ by replacing the parts of $\ell_0\ell_1$ inside each Ω^1_n with unknotted 1-tangles, or we can recover $p(\ell_2)$ by doing the same to each Ω^2_n.

Note that the two families of balls can intersect each other in nontrivial ways. We will retroactively modify $\{\Omega^2_n\}$ (so that it no longer comes from a translation-respecting family of balls for $\ell_0\ell_2$) to control the intersections. Without rehashing the details, given any m we can follow similar reasoning to the proof of Lemma 5.4.1 to pick a new Ω^2_m that, while still disjoint from all other Ω^2_n, is either disjoint from all Ω^1_n or is contained in Ω^1_m. This assignment can be made without unduly enlarging Ω^2_m, so we can inductively modify Ω^2_n as above for all $n \in \mathbb{Z}$ to get a complete new end-properly embedded family of balls.

Now let us consider what happens when we replace that parts of $\ell_0\ell_1$ inside each Ω^2_n with unknots. For n with Ω^2_n contained inside Ω^1_n, the presence of Ω^2_n insider Ω^1_n expresses κ as a connected sum of itself with the unknot. Therefore, unknotting the inside of Ω_n^2 is the same as unknotting the inside of Ω_n^1. For n with Ω_n^2 disjoint from Ω_n^1, let us consider the ball composed as a union of Ω_n^1 with Ω_n^2 and a small regular neighorhood of the strand of $\ell_0\ell_1$ connecting Ω_n^1 with Ω_n^2. This ball contains a connected sum of two copies of κ, and replacing either summand with an unknot gives the same result.

Therefore, unknotting the insides of all of the Ω_n^2 is the same as unknotting the insides of all of the Ω_n^1, so ℓ_2 is isotopic to ℓ_1. □

Theorem 5.4.3. *If $\ell_0 \in A_{\mathbb{Z}}$ has height 0 and $\ell_1 \in A_{\mathbb{Z}}$ has the property that $q(\ell_1)$ is uniquely determined by its image under* Bh*, then $q(\ell_0)q(\ell_1)$ is also uniquely determined by its image under* Bh*.*

Proof. By induction on the number of prime factors of the oriented knot κ with $\kappa^\bullet = \overline{q}(s^{-1}(\ell_0))$, we can assume that κ is prime. The base case is trivial, since if κ is the unknot then $q(\ell_0)$ is the identity for $A_{\mathbb{Z}}/\mathcal{I}_{\mathbb{Z}}$ as a monoid, and the induction step works because the knot-type element of $A_{\mathbb{Z}}/\mathcal{I}_{\mathbb{Z}}$ corresponding to a connected sum of knots is the product of the knot-type elements corresponding to the individual factors.

Now suppose some $\ell \in A_{\mathbb{Z}}$ has $\mathrm{Bh}(q(\ell)) = \mathrm{Bh}(q(\ell_0)q(\ell_1))$, or equivalently $p(\ell) = p(\ell_0\ell_1)$. Then we can apply Lemma 5.4.1 to ℓ with κ as the prime oriented knot to deduce that $\overline{q}(s^{-1}(\ell))$ can be written as $\kappa^\bullet\overline{q}(s^{-1}(\ell_2))$ for some ℓ_2. Therefore, we have $q(\ell) = q(\ell_0)q(\ell_2)$, so we can apply Lemma 5.4.2. This implies that $p(\ell_1) = p(\ell_2)$ and therefore $\mathrm{Bh}(q(\ell_1)) = \mathrm{Bh}(q(\ell_2))$. By our original assumption on ℓ_1, it follows that $q(\ell_1) = q(\ell_2)$ and hence $q(\ell) = q(\ell_0)q(\ell_1)$, as desired. □

5.5 Extremal Spanning Disks

In this chapter, we will use a very specific notion of two spanning disks being isotopic: For a fixed $\overline{\ell} \in \overline{A}$ and two spanning disks $\overline{D}_1, \overline{D}_2$ for $\overline{\ell}$ in $S^1 \times I^2$, an *isotopy from $\overline{D}_1$ and $\overline{D}_2$* is a continuous family of elements of $\overline{\mathcal{J}}$ (all automorphisms of $S^1 \times I^2$ that don't necessarily fix the boundary pointwise), indexed over $[0, 1]$, such that (a) all ambient automorphisms in the family fix $\overline{\ell}$ as a set (but not necessarily pointwise), (b) at all times the result of applying the ambient automorphism to $\overline{D}_1$ is still a valid spanning disk, (c) the ambient automorphism at time 0 is the identity, and (d) the ambient automorphism at time 1 sends $\overline{D}_1$ to $\overline{D}_2$. That is, we allow movement in both the boundaries of the spanning disks and the exact intersection points of the spanning disks with $\overline{\ell}$, but we do not allow $\overline{\ell}$ to move, even temporarily. Whenever an isotopy from $\overline{D}_1$ to $\overline{D}_2$ exists, we will say $\overline{D}_1$ and $\overline{D}_2$ are *isotopic*. Let $\overline{\mathcal{D}}_{\overline{\ell}}$ denote the set of equivalence classes of minimal spanning disks for $\overline{\ell}$ up to isotopy.

For $\ell \in A$ we can define isotopy similarly. For $\ell \in A_{\mathbb{Z}}$, we will let $\mathcal{D}_\ell$ be the equivalence classes of minimal spanning disks for ℓ up to both isotopy and integral translation.

There is a natural relation $\leq$ on $\mathcal{D}_\ell$: For $\Psi_1, \Psi_2 \in \mathcal{D}_\ell$, we will write $\Psi_1 \leq \Psi_2$ if there are representatives $D_1 \in \Psi_1$ and $D_2 \in \Psi_2$ such that (a) D_1 and D_2 are disjoint, (b) the region of $\partial(\mathbb{R} \times I^2)$ bounded by $\partial D_1 \cup \partial D_2$ doesn't contain any points of $\mathbb{Z} \times \{\pm 1\} \times \{0\}$, and (c) D_2 lies on the side of D_1 in which the x values are bounded below.

Lemma 5.5.1. *The relation $\leq$ is a partial order.*

Proof. The proof is essentially the same as that of Lemma 2.14 in [44], using the complement of ℓ as M. We note that because the D_i intersect ℓ minimally, they must be incompressible and are therefore π_1-injective. □

Below we will look into elements of $\mathcal{D}_\ell$ that are extremal under the partial order $\leq$.

In order to reduce the scope of the problem, we check that we need only consider compact regions of $\mathbb{R} \times I^2$:

Lemma 5.5.2. *For all $\ell \in A_{\mathbb{Z}}$, there is a compact region of $\mathbb{R} \times I^2$ such that every element of $\mathcal{D}_\ell$ has a representative in the region.*

Proof. For a given $\ell \in A_{\mathbb{Z}}$, pick a fixed minimal spanning disk $\overline{D}$ for $s^{-1}(\ell) \in \overline{A}$, and consider the preimage Δ of $\overline{D}$ in $\mathbb{R} \times I^2$. Then as in the proof of Theorem 5.4.3, Δ is a union of pairwise-disjoint minimal spanning disks Δ_i for ℓ, all of which are $\mathbb{Z}$-translates of each other.

We will show that every element of $\mathcal{D}_\ell$ has a representative contained in the region of $\mathbb{R} \times I^2$ bounded between Δ_{-k} and Δ_k, where $k = (h(\ell) + 2)^2$.

Consider a fixed $\Psi \in \mathcal{D}_\ell$. By Theorem 3.3.1, all representatives of all elements of $\mathcal{D}_\ell$ have the same algebraic intersection number with the strands of ℓ, and therefore we can choose a representative $D \in \Psi$ with the same boundary as Δ_0. We also can insist for D to be in general position, thereby intersecting Δ transversely and away from ℓ, and take it to minimize the number of components of $D \cap \Delta$.

Suppose that D intersects Δ_{-k} or Δ_k. Without loss of generality, let us assume that it intersects Δ_k. Let C_k be an innermost (on Δ_k) circle of intersection between D and Δ_k. Then the annulus on D on the outside of C_k must intersect Δ_i for each $i = 1, \ldots, k-1$, and at least one of the circles of intersection is not nullhomologous in the aforementioned annulus. For each such i, let C_i be an innermost (on Δ_i) circle of intersection among those that are not nullhomologous in the annulus. For each $i = 1, \ldots, k$, let E_i and E'_i be the disks on Δ_i and D, respectively, bounded by C_i. By Lemma 3.1.2, the number of intersection points of ℓ with E_i and E'_i must be equal for each i. It must be that E_k intersects ℓ at least once, or else E'_k would not either, and together they would form a sphere through which we could push D to reduce the number of intersections with Δ. Therefore, each E_i has at least one intersection with ℓ. For each i, let $n_i \in \mathbb{Z}$ be the index of an arbitrary component of ℓ intersecting E_i. For i with $h(\ell) \leq i \leq k - h(\ell)$, the component n_i of ℓ must also intersect D, because otherwise it would have to intersect Δ more than $h(\ell)$ times to "escape" the annulus. Furthermore, for $i, j \in \mathbb{Z}$ with $|i - j| \geq h(\ell)$, n_i and n_j must be distinct, because no single component of ℓ can intersect Δ_i and Δ_j without also intersecting the $|i-j|-1$ components of Δ in between. Hence, by our choice of k, at least $h(\ell)+1$ distinct components of ℓ intersect D, so the total number of intersections of ℓ with D is greater than $h(\ell)$, a contradiction.

Therefore, we have shown that every element of $\mathcal{D}_\ell$ has a representative contained in the compact region of $\mathbb{R} \times I^2$ between Δ_{-k} and Δ_k. □

Next we prove that, under certain conditions, any pair of elements in the partially ordered set have a mutual greater element and mutual lesser element:

Lemma 5.5.3. *For $\ell \in A_\mathbb{Z}$ with $h_-(\ell) = 0$ or $h_+(\ell) = 0$, given two elements Ψ_1, Ψ_2 of $\mathcal{D}_\ell$, there are other elements Ψ_- and Ψ_+ with $\Psi_- \leq \Psi_1 \leq \Psi_+$ and $\Psi_- \leq \Psi_2 \leq \Psi_+$.*

Proof. First we will assume without loss of generality that $h_-(\ell) = 0$.

For such Ψ_1 and Ψ_2, let us pick D_1 and D_2 representing Ψ_1 and Ψ_2 in such a way that (a) $\partial D_1 = \partial D_2$, (b) D_1 and D_2 intersect transversely (including on their boundaries) away from ℓ, and (c) the number of components of intersection is minimal.

Consider the union $D_1 \cup D_2$. The complement of $D_1 \cup D_2$ in $\mathbb{R} \times I^2$ has two unbounded components and some positive number of bounded components. Let $R_+ = R_+(D_1 \cup D_2)$ denote the closure of the component with the x coordinates unbounded above, and $R_- = R_-(D_1 \cup D_2)$ the closure of the component with the x components unbounded below. Without loss of generality, we can assume that the components of $D_2 \cap R_+$ occur away from $\partial D_1 = \partial D_2$, since otherwise we could make the assumption true by relabelling Ψ_1 and Ψ_2. Note that R_+ is a subspace of $R_+(D_1)$, which has no interesting topology of its own.

Now let us suppose that ∂R_+ has positive genus. Then $D_2 \cap R_+$ is compressible as a surface in $R_+(D_1)$. Let C_0 be a compressing circle on $D_2 \cap R_+$. Then by the definition of compressibility, the disk E_0 in D_2 bounded by C_0 is not contained entirely in R_+. Now look at the circles of intersection between E_0 and D_1, and among the circles that are outermost in D_2, pick one that is outermost in D_1 and call it C. (This C will not necessarily be outermost among all circles of intersection between E_0 and D_1, only among those that are outermost in D_2.) Let E_1 and E_2, respectively, be the disks in D_1 and D_2 bounded by C. Both E_1 and E_2 have natural orientations inherited from D_1 and D_2. They each therefore induce an orientation on C, but our choice of C ensures that the two orientations on C are opposite. In other words, $E_1 \cup E_2$ is an immersed sphere in $\mathbb{R} \times I^2$ with a well-defined orientation. It is nullhomologous, and therefore its net intersection with ℓ is 0. Because $H_-(\ell, D_1) = H_-(\ell, D_2) = 0$, there are no negative intersections of ℓ with E_1 or E_2, and therefore neither E_1 nor E_2 has any intersections with ℓ of any sign. Let us find a circle of intersection between D_2 and E_1 that is innermost on E_1. (This innermost circle might just be $\partial E_1 = C$). Then the circle encloses disks E_1' and E_2' on D_1 and D_2, respectively. Because E_1 is disjoint from ℓ, so is E_1'. Also, E_1' is disjoint from E_2', because the boundary of E_1' was innermost in E_1. Therefore, $E_1' \cup E_2'$ is an embedded sphere in $\mathbb{R} \times I^2$, which is not necessarily oriented in a consistent way. The inside of the sphere cannot include any part of ℓ, since the part would have to enter and leave by E_2', but E_2' cannot have any negative intersections with ℓ. Therefore, we can push D_2 through the ball to reduce the number of intersections between D_1 and D_2, obtaining a contradiction.

We have concluded that ∂R_+ has zero genus, and the same reasoning applies to ∂R_-. Let $D_\pm$ be spanning disks formed by pushing $R_\pm \cap (D_1 \cup D_2)$ off of $D_1 \cup D_2$.

It remains to show that $D_\pm$ realize the height of ℓ. Whenever ℓ intersects D_1 or D_2, it does so positively. Since R_+ is on the positive/"right" side of each D_i, the parts of each D_i contained in ∂R_+ are oriented so that positive intersections point into R_+. Therefore, the intersections between ℓ and D_+ are positive, and by the same reasoning we can say the same of ℓ and D_-. Therefore $H_-(\ell, D_\pm) = 0$. Since each $D_\pm$ has essentially the same boundary as D_1 and D_2, they both give the same spanning disk framing to ℓ as D_1 and D_2. Therefore, $H_+(\ell, D_\pm) = H_+(\ell, D_1) = H_+(\ell, D_2) = h_+(\ell)$. This shows that $D_\pm$ are both minimal. Therefore they represent classes Ψ_+ and Ψ_- with the desired properties. □

Lemma 5.5.4. *For $\ell \in A_{\mathbb{Z}}$ with $h_+(\ell) = h_-(\ell) = 1$, if ℓ is knotless then given two elements Ψ_1, Ψ_2 of $\mathcal{D}_\ell$, there are other elements Ψ_- and Ψ_+ with $\Psi_- \leq \Psi_1 \leq \Psi_+$ and $\Psi_- \leq \Psi_2 \leq \Psi_+$.*

Proof. We begin as in the proof of Lemma 5.5.3, fixing D_1, D_2, R_+, and R_-, and supposing for the sake of contradiction that ∂R_+ or ∂R_- has positive genus.

No circle of intersection between D_1 and D_2 can enclose a disk on either D_i that is disjoint from ℓ: Supposing that such a disk E on some D_i exists, we could find an innermost circle of intersection on E, and the disk E' inside that innermost circle would also be disjoint from ℓ. That new disk E' would also be disjoint from the disk E'' in the other D_i with the same boundary as E'. By the fact that both spanning disks in question are minimal, E'' is also disjoint from ℓ. Then we could push one spanning disk through the ball enclosed by $E' \cup E''$, contradicting the minimal intersection assumption.

By similar reasoning, we can see that no circle of intersection between D_1 and D_2 can enclose a disk on either D_i that intersects ℓ exactly once: Supposing that such a disk E on some D_i exists, the circles of intersection between D_1 and D_2 on E must all be nested, because none of them bounds a disk in E disjoint from ℓ. Therefore there is an innermost disk E' that meets ℓ exactly once. If E'' is the disk in the other D_i with the same boundary as E', then $E' \cup E''$ is an embedded sphere. By the fact that both spanning disks in question are minimal, E'' also intersects ℓ exactly once. Therefore the ball inside $E' \cup E''$ encloses a 1-tangle. By the assumption that ℓ is knotless, the 1-tangle is trivial, which again implies that we could cancel an intersection between D_1 and D_2 by pushing, a contradiction.

Since D_1 and D_2 each intersect ℓ only twice, the previous two paragraphs imply that for each i, each component of $D_1 \cap D_2$ as a subset of D_i encircles both points of $D_i \cap \ell$. Therefore, the intersection circles are nested in each D_i. This means that each component of $D_1 \cap \partial R_+$, $D_2 \cap \partial R_+$, $D_1 \cap \partial R_-$, or $D_2 \cap \partial R_-$ is either a disk or an annulus. Thus the whole surfaces ∂R_+ and ∂R_- are composed by gluing together disks and annuli (including the unbounded components) along their boundaries, so $\partial R_\pm$ cannot have nonzero genus. Furthermore, in the decompositions of ∂R_+ and ∂R_- into annuli and disks, each surface only has one disk piece.

Now, as before, let $D_\pm$ be the spanning disks formed by pushing $R_\pm \cap (D_1 \cup D_2)$ off of $D_1 \cup D_2$. Because of the nesting of the components of $D_1 \cap D_2$ on each of D_1 and D_2, the components of each $D_i \cap \partial R_\pm$ that are annuli do not intersect ℓ at all, and those that are disks intersect ℓ twice. Therefore, D_1 and D_2 each intersect ℓ exactly twice, so they are minimal. □

We now have sufficient conditions for $\mathcal{D}_\ell$ to have extremal elements.

Lemma 5.5.5. *For $\ell \in A_{\mathbb{Z}}$ with either (a) $h_-(\ell) = 0$, (b) $h_+(\ell) = 0$, or (c) ℓ is knotless and $h_+(\ell) = h_-(\ell) = 1$, $\mathcal{D}_\ell$ has a greatest element and a least element.*

Proof. Suppose we have some ℓ as above. The set $\mathcal{D}_\ell$ must be at most countable, because each element will have a representative that is piecewise-linear with vertices at points with rational coefficients, as any spanning disk can be approximated in that way. Choose a counting function $\mathbb{N} \to \mathcal{D}_\ell$ (possibly with repeats), denoted $i \mapsto \Psi_i$. Fix Δ_0, Δ_{-k}, and Δ_k as in Lemma 5.5.2.

We will recursively define two sequences $\{D_i\}$ and $\{D'_i\}$ of minimal spanning disks for ℓ, all of which will be between Δ_{-k} and Δ_k. First, let D_1 be a representative for Ψ_1, chosen to coincide with Δ_0 on its boundary and to sit between Δ_{-k} and Δ_k as provided by Lemma 5.5.2. Let D'_1 be equal to D_1.

Now, having defined D_i and D'_i for some i, we will define D_{i+1} and D'_{i+1} as follows. Let $\Psi'_i \in \mathcal{D}_\ell$ be the class of D'_i. Now, depending on the signed heights of ℓ, we will apply Lemma 5.5.3 or Lemma 5.5.4 to Ψ'_i and Ψ_{i+1}. Specifically, we will use D'_i as our representative of Ψ'_i, and for our representative of Ψ_{i+1} we will pick an element D_{i+1} that, together with D'_i, fits the description at the start of the proofs of Lemma 5.5.3 and Lemma 5.5.4. The proof of Lemma 5.5.2 shows that even under those restrictions, we can take D_{i+1} to sit between Δ_{-k} and Δ_k. That allows us to define D'_{i+1} to be a a minimal spanning disk lying slightly to the right of D'_i and D_{i+1}. This D'_{i+1} will still be in the region between Δ_{-k} and Δ_k.

We now have a sequence $\{D'_i\}_{i\in\mathbb{N}}$ of mutually disjoint minimal spanning disks for ℓ, all with isotopic boundaries and all contained between Δ_{-k} and Δ_k. We can now apply Kneser–Haken finiteness as in Theorem 1 of [20]. Let M be the part of the complement of ℓ in $\mathbb{R} \times I^2$ that lies between Δ_{-k} and Δ_k, and consider the parts of the D'_i in M. Since the D'_i are minimal spanning disks, the resulting surfaces are incompressible and all have identical Betti numbers. The theorem ensures that any large subcollection of the D'_i includes two that are parallel and hence isotopic. Therefore, there are only finitely many $\Psi \in \mathcal{D}_\ell$ represented by the D'_i. This implies in turn that there is a $\Psi_+ \in \mathcal{D}_\ell$ represented by infinitely many D'_i. This Ψ_+ will then satisfy $\Psi_i \leq \Psi_+$ for all i. Therefore, Ψ_+ is a greatest element for $\mathcal{D}_\ell$.

By the same reasoning, $\mathcal{D}_\ell$ also has a least element Ψ_-. □

The last result in this section will be used later for controlling the potential discrepancy between the diversity of minimal spanning disks of an annular 1-tangle and the (potentially larger) diversity of minimal spanning disks for the corresponding $\mathbb{Z}$-tangle.

Lemma 5.5.6. *For $\ell \in A_{\mathbb{Z}}$ satisfying the hypotheses of Lemma 5.5.5, the greatest and least elements of $\mathcal{D}_\ell$ each have representatives that are disjoint from their $\mathbb{Z}$-translates.*

Proof. Let Ψ_+ denote the greatest element of $\mathcal{D}_\ell$. We will focus on Ψ_+, since the proof for the least element is equivalent.

Consider a minimal spanning disk for $s^{-1}(\ell)$ and let Δ be its preimage in $\mathbb{R} \times I^2$. As usual, denote the components of Δ by Δ_i, with the indices $i \in \mathbb{Z}$ reflecting the relative positions of the components. Let D be a representative of Ψ_+.

If it is possible to choose Δ and D in such a way that D is disjoint from Δ, then the desired property applies to D. Therefore, suppose that any choice of Δ and D intersect.

Let us choose Δ and D in such a way that (a) the intersections between D and Δ are transverse and occur away from ℓ, (b) the boundary of D almost coincides with that of Δ_0, but is pushed slightly in the negative-x direction, (c) the maximum m of all n such that D intersects Δ_n is minimized (subject to the previous conditions), and (d) the number of intersections between D and Δ_m is minimized (subject to the previous conditions).

Suppose that $m \geq 0$. By the reasoning in the proofs of Lemma 5.5.3 and Lemma 5.5.4 applied to D and Δ_m, we find another minimal spanning disk Δ'_m that sits slightly to the right of $D \cup \Delta_m$. It is contained in the space between Δ_m and Δ_{m+1}, so it is disjoint from all of its integral translates. In particular, the disk Δ'_0 obtained by translating Δ'_m by $-m$ have boundary just to the right of Δ_0. Therefore the union of all Δ'_n would provide a better alternative choice for Δ, since D is disjoint from Δ'_m, so we have a contradiction. Therefore, $m = -1$, since if D intersects Δ without intersecting Δ_0 it must intersect Δ_{-1}.

Now let us apply the constructions in Lemma 5.5.3 and Lemma 5.5.4 to D and Δ_{-1}. This gives us a minimal spanning disk for ℓ with nearly the same boundary as D (lying to the left of $\partial\Delta_0$) but that is disjoint from D and from Δ. By the assumption that D represents the greatest element of $\mathcal{D}_\ell$, the new disk is isotopic to D. This contradicts our supposition that D could not be chosen to be disjoint from Δ. Therefore, we are done. □

5.6 Height Conditions for Injectivity

This section will show that any pair of $\mathbb{Z}$-tangles constituting a counterexample to injectivity of the Birman–Hilden map must have total height at least 3 and positive- and negative heights each at least 1.

Theorem 5.6.1. *If $\ell \in A_{\mathbb{Z}}$ has $h_-(\ell) = 0$ or $h_+(\ell) = 0$, then $q(\ell)$ is uniquely determined by its image under* Bh.

Proof. Suppose $\ell' \in A_{\mathbb{Z}}$ is another translation-invariant $\mathbb{Z}$-tangle with $p(\ell') = p(\ell)$. Then by Lemma 3.1.3, ℓ' has the same signed heights as ℓ.

Since ℓ and ℓ' are isotopic, there is a bijection between $\mathcal{D}_\ell$ and $\mathcal{D}_{\ell'}$ that respects the partial orders. By Lemma 5.5.5, $\mathcal{D}_\ell$ and $\mathcal{D}_{\ell'}$ have greatest elements. The maxima must be mapped to each other under the aforementioned bijection. By Lemma 5.5.6, we can choose representatives D_0 and D'_0 for the maximal classes such that D_0 and D'_0 each are disjoint from their integral translates. By translating D'_0 and then adjusting it in a neighborhood of its boundary without interfering with ℓ', we can also require for simplicity that $\partial D_0 = \partial D'_0$. Then there is a boundary-preserving isotopy $\iota_0 \in \mathcal{I}$ that sends ℓ' to ℓ and D'_0 to D_0.

For every $n \in \mathbb{N}$, let D_n and D'_n be the translates of D_0 and D'_0, respectively, by n in the x direction. These new spanning disks, like D_0 and D'_0, represent the maxima of $\mathcal{D}_\ell$ and $\mathcal{D}_{\ell'}$. Therefore there is an isotopy $\iota_n = \tau^n \iota_0 \tau^{-n}$ sending D'_n to D_n while sending ℓ' to ℓ. We cannot assume that ι_n is equal to ι_0.

Since D'_1 is a representative of the maximum of $\mathcal{D}_{\ell'}$, $\iota_0(D'_1)$ will represent the maximum of $\mathcal{D}_\ell$. Therefore it is isotopic to D_1 in the sense used to define $\mathcal{D}$. By retroactively modifying ι_0 we can ensure that $\iota_0(D'_1)$ can be isotoped to D_1 without moving any points of intersection with ℓ. Still, this isotopy from $\iota_0(D'_1)$ to D_1 might see the disk temporarily intersecting D_0. Consider that D_0, thought of as a punctured surface in the complement of ℓ in $\mathbb{R} \times I^1$, is incompressible because it minimizes height and therefore it is π_1-injective. Then we can use basic obstruction theory to find a homotopy from $\iota_0(D'_1)$ to D_1 contained entirely in $R_+(D_0)$, since the triviality of the higher homotopy groups of $R_-(D_0)$ allows us to push homotopies of cells of dimension 1 and 2 back through D_0. By Corollary 5.5 of [50], this homotopy can be promoted to an ambient boundary-preserving isotopy of $R_+(D_0)$. Composing this isotopy with ι_0, we obtain a boundary-preserving isotopy ι of $\mathbb{R} \times I^2$ that sends ℓ' to ℓ, D'_0 to D_0, and D'_1 to D_1. Therefore ι maps $R(D'_0, D'_1)$ onto $R(D_0, D_1)$, where $R(D_0, D_1) = R_+(D_0) \cap R_-(D_1)$ and so on.

There are, of course, natural identifications $D_0 \cong D_1$ and $D'_0 \cong D'_1$ by translation. The isotopy ι induces its own identifications $D'_0 \cong D_0$ and $D'_1 \cong D_1$. Let b be the composition

$$D_0 \xrightarrow{\tau} D_1 \xrightarrow{\iota^{-1}} D'_1 \xrightarrow{\tau^{-1}} D'_0 \xrightarrow{\iota} D_0.$$

The restriction of b to $D_0 \cap \ell$ is the identity. Therefore, b can be thought of as a pure braid with $h(\ell)$ strands, depending on an exact identification of D_0 with a standard $h(\ell)$-punctured disk. We will first show by contradiction that b cannot represent a nontrivial braid, and then use the fact that b represents the trivial braid to obtain $q(\ell') = q(\ell)$.

Case 1: Nontrivial braid.

For this part we will assume without loss of generality that $h_+(\ell) = 0$.

Suppose for the sake of contradiction that the braid represented by b is nontrivial.

The points of $\ell \cap D_1$ can naturally be indexed with $\{1, \ldots, h(\ell)\}$: Each corresponds uniquely to a point of $s^{-1}(\ell) \cap \overline{D}$, where $\overline{D}$ is the spanning disk for $s^{-1}(\ell)$ corresponding to D_1, and the points of $s^{-1}(\ell) \cap \overline{D}$ are indexable by the order in which they appear on $s^{-1}(\ell)$.

Let ℓ_b denote the tangle of infinitely many strands in $R_+(D_0)$ obtained (up to boundary-fixing isotopy) by, for all $n \in \mathbb{N}$, inserting a copy of b in braid form into a neighborhood of D_n.[1] Because b equals the discrepancy between the restrictions of $\tau^{n-1}\iota\tau^{1-n}$ and $\tau^n \iota \tau^{-n}$ to D_n for each n, the isotopies $\tau^n \iota \tau^{-n}$ together define an identification of the part of ℓ' in $R_+(D'_0)$ with ℓ_b. Together with the previous identification, this implies that there is a diffeomorphism from $R_+(D_0)$ to itself that sends ℓ_b to ℓ and restricts to the identity on the boundary.

[1] Our precise convention for the method of insertion will not matter.

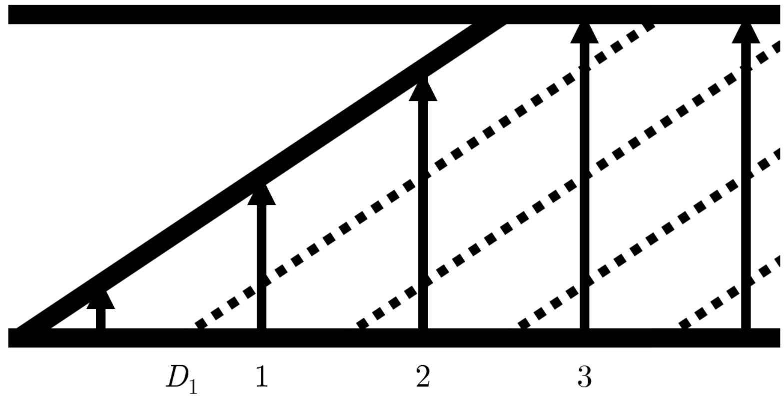

Figure 5.1: An illustration of $R_+(D_0)$ in a case where $h(\ell) = 3$. The dotted lines represent spanning disks D_n for $n > 0$ and the vertical lines represent strands of ℓ. We have labelled each strand that intersect D_1 according to the first n for which the strand is included in $T_n(\ell)$.

For n in $\{1, \ldots, h(\ell)\}$, let $T_n(\ell)$ be the n-tangle in $R_+(D_0)$ where the n strands are the components of $\ell \cap R_+(D_0)$ that include the points of $\ell \cap D_1$ with indices $1, \ldots, n$. Similarly, let $T_n(\ell_b)$ be the n-tangle in $R_+(D_0)$ with n components chosen from ℓ_b in the same way.

Now, for n in $\{1, \ldots, h(\ell)\}$, let b_n be the pure n-strand braid obtained from b by considering only the first n strands. Let m be the least n such that b_n is nontrivial. Notice that each $T_m(\ell_b)$ differs from $T_m(\ell)$ by the insertion of m braids, specifically, the braids $b_1, \ldots, b_m$ at the positions of $D_m, \ldots, D_1$, respectively. By the way that we chose m, it follows that $T_m(\ell_b)$ and $T_m(\ell)$ differ only by the insertion of b_m. However, we have also determined that $\ell_b \cap R_+(D_0)$ and $\ell \cap R_+(D_0)$ are isotopic, which implies that $T_m(\ell_b)$ and $T_m(\ell)$ are isotopic. Then Theorem 4.1.1 implies that b_m is trivial. This is a contradiction, so we can move on to Case 2.

Case 2: Trivial braid.

Since b represents the trivial braid, we can adjust ι retroactively to make it so that b is the identity on D_0. Since ι maps $R(D'_0, D'_1)$ to $R(D_0, D_1)$, and those are both fundamental domains for translation of $\mathbb{R} \times I^2$, ι lifts to a unique corresponding isotopy $\bar{\iota} \in \overline{\mathcal{I}}$, which will send $s^{-1}(\ell')$ to $s^{-1}(\ell)$. Therefore, we have $\overline{q}(s^{-1}(\ell')) = \overline{q}(s^{-1}(\ell))$ and hence also $q(\ell') = q(\ell)$. □

Theorem 5.6.2. *If $\ell \in A_{\mathbb{Z}}$ has $h(\ell) \leq 2$, then $q(\ell)$ is uniquely determined by its image under* Bh.

Proof. By Theorem 5.6.1, we can assume that $h_+ = h_- = 1$, and by Theorem 5.4.3, we can assume that ℓ is knotless.

Now we can largely repeat the proof of Theorem 5.6.1. All parts of the proof carry over without modification except for Case 1, so let us enter that case in our new context.

Under our supposition that b represents a nontrivial braid on two strands, it is simply the m^{th} power of a full twist for some nonzero $m \in \mathbb{Z}$.

Let us look at the 2^{nd} translates D_2 and D_2' of D_0 and D_0'. Applying to D_2 and D_2' the same logic that was applied to D_1 and D_1', there is an isotopy of $\mathbb{R} \times I^2$ that sends ℓ' to ℓ, D_0' to D_0, and D_2' to D_2. Therefore it maps the 4-stranded tangle $\ell' \cap R(D_0', D_2')$ in $R(D_0', D_2')$ to the 4-tangle $\ell \cap R(D_0, D_2)$ in $R_{(}D_0, D_2)$. Like in the proof of Theorem 5.6.1, let ℓ_b be the 4-tangle in $R(D_0, D_2)$ obtained (up to boundary-fixing isotopy) by inserting a copy of b in braid form into a neighborhood of D_1. Again as in the previous proof, ι and $\tau\iota\tau^{-1}$ together define an identification of ℓ' as a tangle in $R(D_0', D_2')$ with ℓ_b as a tangle in $R(D_0, D_2)$. Together with the previous identification, this implies there is a diffeomorphism from $R(D_0, D_2)$ to itself that sends ℓ_b to ℓ and restricts to the identity on $\partial R(D_0, D_2) \setminus D_2$. In other words, there is a braid b' on D_2 such that $\ell^{b'}$ is isotopic to ℓ_b, where $\ell^{b'}$ is the result of inserting b' into ℓ at D_2.

As an intermediate step, we will show that b' is trivial: For each of ℓ, ℓ_b, and $\ell^{b'}$ as 4-tangles in $R_{(}D_0, D_2)$, let us consider the 2-tangle in $R_{(}D_0, D_2)$ formed just by the the two strands contacting D_2. The first two 2-tangles must be isotopic because the strands in question avoid D_1, and the second and third 2-tangles are isotopic because ℓ_b and $\ell^{b'}$ are isotopic. Therefore, by transitivity, the 2-tangles obtained from ℓ and $\ell^{b'}$ are isotopic. Those two differ by applying b' to D_2. Since the two strands each contact D_2 only once each, we can apply Theorem 4.1.1, obtaining the triviality of b'.

Now we know that ℓ and ℓ_b, as tangles in $R_{(}D_0, D_2)$, are isotopic. Note that we can actually obtain ℓ_b up to isotopy just by knowing $\ell \subset R(D_0, D_2)$ and m, without D_1, by applying a three-dimensional Dehn twist of order m in the sense of [40] to ℓ. Though we cannot apply Theorem 4.1.2 directly, as the part of ℓ in $R(D_0, D_2)$ has the wrong number of components, the proof carries over with minimal modification. Hence there is a disk in $R(D_0, D_2)$ with the same boundary as D_1 and with no points of intersection with ℓ. This disk, taken as a spanning disk in $\mathbb{R} \times I^2$, violates our previous conclusion of $h(\ell) = 2$. Therefore, we have a contradiction, so we can go on to Case 2. $\square$

Chapter 6

Points of Comparison

6.1 The Image of the Birman–Hilden Map

In this section we will show that $p(A_{\mathbb{Z}})$ is a proper subset of $(A/\mathcal{I})_{\mathbb{Z}}$, and hence the inclusion in Equation (2.3) is not surjective. Let μ be the $\mathbb{Z}$-tangle shown in Figure 6.1.

Theorem 6.1.1. *The class $p(\mu)$ is in $(A/\mathcal{I})_{\mathbb{Z}}$ but not $p(A_{\mathbb{Z}})$.*

Proof. It is easy to see that μ is isotopic to $\tau\mu$, and hence $p(\mu)$ is translation-invariant.

Now suppose that there is a $\ell \in A_{\mathbb{Z}}$ isotopic to μ. (A natural candidate for such an isotopy would be some kind of "limit"

$$\cdots \circ (\tau^3 \iota \tau^{-3}) \circ (\tau^2 \iota \tau^{-2}) \circ (\tau \iota \tau^{-1}) \circ \iota,$$

where ι is one of the isotopies sending μ to $\tau\mu$, but even if ι is chosen to be compactly-supported there will still be points in $\mathbb{R} \times \mathcal{I}^2$ whose images under the above sequence of composed isotopies will not stabilize.)

Since μ has height 0, ℓ must also have height 0. Since $p(\mu)/\infty$ is unknotted, $q(s^{-1}(\ell))$ must be the unknotted annular 1-tangle. Therefore, we can take ℓ to be the trivial $\mathbb{Z}$-tangle.

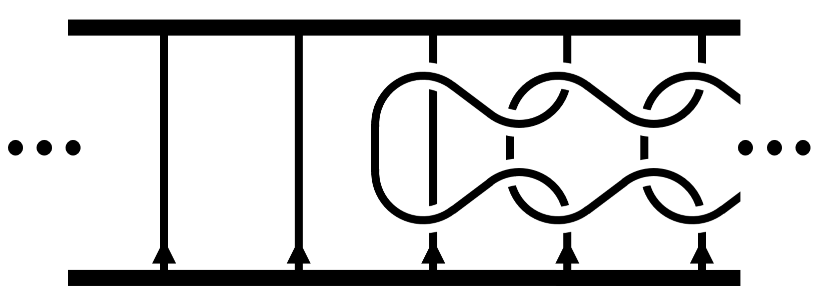

Figure 6.1: A $\mathbb{Z}$-tangle that is translation-invariant up to isotopy but not isotopic to any translation-invariant $\mathbb{Z}$-tangle.

Consider the spanning disks for ℓ of the form $\{x\} \times I^2$ for $x \in \mathbb{Z} + \frac{1}{2}$. All such disks have height 0 and together they constistute a biproper embedding $\mathbb{Z} \times I^2 \to \mathbb{R} \times I^2$. Let $\iota \in \mathcal{I}$ be a specific isotopy with $\iota \circ \ell = \mu$, and let us consider the composition $\mathbb{Z} \times I^2 \to \mathbb{R} \times I^2 \to \mathbb{R} \times I^2$. This is again a biproper embedding consisting of many spanning disks of height 0.

By properness, there is a spanning disk of height 0 contained in the tangled repeating part of μ. However, the same disk could then be used for the translation-invariant $\mathbb{Z}$-tangle formed using just that repeating portion. That would then imply that a product of two bifoils would have height 0, which is a contradiction. Hence, we are done. □

Taking the problem one step further, one could also ask whether $p(A_\mathbb{Z})$ and $(A/\mathcal{I})_\mathbb{Z}$ even have the same cardinality. Since $q(A_\mathbb{Z})$ is countable, $p(A_\mathbb{Z})$ is as well. We conjecture that $(A/\mathcal{I})_\mathbb{Z}$ is uncountable, but we do not yet have a proof of this.

6.2 Annular Knots

The theory of annular 1-tangles can readily be compared to the similar theory of nullhomologous annular knots (see, for example, [52]). The Birman–Hilden map has natural equivalents in this other context; in this case the function takes values in the set of isotopy classes of $\mathbb{Z}$-*links*, that is, end-proper embeddings $S^1 \times I \to \mathbb{R} \times I^2$ for which $x \to \infty$ as $n \to \infty$ and $x \to -\infty$ as $n \to -\infty$.

The Birman–Hilden map for nullhomologous annular knots is not injective. A counterexample is shown in Figure 6.2: One can see that the two annular knots are distinct by embedding the thickened annulus into $\mathbb{R}^3$ as shown; one knot becomes unknotted under this embedding and the other becomes a trefoil. Meanwhile, the corresponding $\mathbb{Z}$-links are isotopic via twisting by varying angles at each fiber $\{\mathrm{pt}\} \times I^2$. Note that this isotopy can be performed in finite time with a well-defined image for each point.

6.3 Weak Isotopy

The example μ in Section 6.1 suggest that alongside isotopy we ought to consider several weaker notions of equivalence between $\mathbb{Z}$-tangles. Three such relations are as follows:

Embedding

Definition 6.3.1. A $\mathbb{Z}$-tangle ℓ *embeds in* another $\mathbb{Z}$-tangle ℓ' if there is an embedding $\psi\colon \mathbb{R} \times I^2 \hookrightarrow \mathbb{R} \times I^2$ such that (a) ψ restricts to the identity on $\partial(\mathbb{R} \times I^2)$ and (b) $\psi \circ \ell = \ell'$. This embedding is not required to be end-proper.

Of course, if ℓ and ℓ' are isotopic then they it follows immediately that they embed in each other.

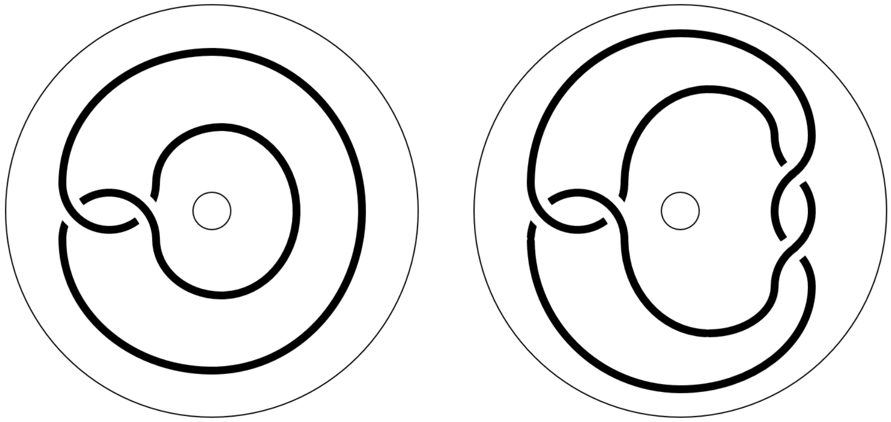

Figure 6.2: Two annular knots. These are a counterexample to injectivity of the version of the Birman–Hilden map for nullhomologous annular knots.

The trivial $\mathbb{Z}$-tangle embeds in μ. To see how, let us first fix a compactly-supported isotopy ι sending μ to $\tau\mu$. Then we will define an embedding ψ of the trivial $\mathbb{Z}$-tangle to μ by defining its restrictions to $(-\infty, x_0] \times I^2$ for each $x_0 \in \mathbb{R}$. Given x_0, let us pick a $n \in \mathbb{N}$ such that $\tau^n \iota \tau^{-n}$ fixes all points in $(-\infty, x_0] \times I^2$. Now define $\psi\colon (-\infty, x_0] \times I^2$ to be

$$\iota^{-1} \circ (\tau\iota^{-1}\tau^{-1}) \circ \cdots \circ (\tau^n \iota^{-1} \tau^{-n}) \circ (\text{inclusion } (-\infty, x_0] \times I^2 \hookrightarrow \mathbb{R} \times I^2).$$

This definition is consistent over all choices of x_0, so there is a well-defined total map ψ on $\mathbb{R} \times I^2$ satisfying our requirements. The image of ψ will necessarily be a proper subset of $\mathbb{R} \times I^2$ and the complement will be a fractal. Depending on the exact choice of ι, the complement might locally look light the product of a Cantor set with $\mathbb{R}^2$.

Conversely, μ does not embed in the trivial $\mathbb{Z}$-tangle: As in the proof in Section 6.1, the existence of such an embedding would contradict the existence of a proper family of spanning disk with height 0 for the trivial $\mathbb{Z}$-tangle. Therefore, embedding is not symmetric.

Conjecturally, the embedding relation should be antisymmetric (on the level of isotopy classes), that is, if two $\mathbb{Z}$-tangles each embed in each other then they must be isotopic.

Weak Isotopy

The next alternative version of isotopy for $\mathbb{Z}$-tangles is perhaps the most visually intuitive.

Definition 6.3.2. A *weak isotopy* of $\mathbb{Z}$-tangles a smooth biproper map $\mathbb{Z} \times I \times [0,1] \to \mathbb{R} \times I^2$ for which the restriction to each time slice $\mathbb{Z} \times I \times \{\text{pt}\}$ defines a $\mathbb{Z}$-tangle. Two $\mathbb{Z}$-tangles are *weakly isotopic* if there is a weak isotopy that restricts to both the two $\mathbb{Z}$-tangles at times 0 and 1.

The smoothness and properness requirements above, which apply to the map as a whole and not just each intermediate $\mathbb{Z}$-tangle, together ensure that a weak isotopy "looks like an isotopy" to an observer who only watches a compact region of $\mathbb{R} \times I^2$.

Unlike embedding, weak isotopy is an equivalence relation. If ℓ and ℓ' are $\mathbb{Z}$-tangles for which ℓ embeds in ℓ', then ℓ and ℓ' are weakly isotopic. Indeed, given ψ realizing the embedding, we can construct a weak isotopy from ℓ' to ℓ by finding a smooth family of embeddings $\mathbb{R} \times I^2 \hookrightarrow \mathbb{R} \times I^2$ interpolating between ψ and the identity. Such a smooth family can be constructed by first finding an isotopy from $\psi(\{0\} \times I^2)$ to $\{0\} \times I^2$ and composing this isotopy with ψ, and then progressively working outward to map each space $\psi([-x_0, x_0] \times I^2)$ back to $[-x_0, x_0] \times I^2$.

Weak isotopy classes of $\mathbb{Z}$-tangles can be very diverse. By the reasoning above, the example μ is weakly isotopic to the trivial $\mathbb{Z}$-tangle by "unzipping" to the right. Similarly, by zipping to the left, it is also weakly isotopic to a translation-invariant $\mathbb{Z}$-tangle of height 2 corresponding to the product of two bifoils. Furthermore, the same logic implies that this weak isotopy class also contains the $\mathbb{Z}$-tangle in Figure 2.6, which is like the aforementioned product except with a full twist incorporated into each zipper loop. These last two $\mathbb{Z}$-tangles fail to be (strongly) isotopic to each other, as demonstrated by Theorem 5.6.2. Meanwhile, are also weakly isotopic to $\mathbb{Z}$-tangles that are not translation-invariant even up to isotopy, since when zipping we can choose a number of twists independently for each individual zip as it is made. The logic of Theorem 5.6.2 implies that, in the family of height-2 $\mathbb{Z}$-tangles obtainable from the example of Figure 2.6 by twisting at each position according to some arbitrary function $\mathbb{Z} \to \mathbb{Z}$, no two members of the family coming from distinct functions $\mathbb{Z} \to \mathbb{Z}$ are isotopic. Thus the weak isotopy class includes uncountably many isotopy classes.

Two weakly isotopic $\mathbb{Z}$-tangles need not be related to each other by embedding in either direction. However, we can also compare weak isotopy to the equivalence relation generated by embeddings. Evidently, given any finite sequence of $\mathbb{Z}$-tangles for which any adjacent pair admits one embedding in the other, all of the $\mathbb{Z}$-tangles in the sequence are weakly isotopic. It is not yet clear whether or not, conversely, any two weakly isotopic $\mathbb{Z}$-tangles are related by the equivalence relation generated by embeddings.

Finite Isotopy

Lastly, we can compare $\mathbb{Z}$-tangles in terms of their relationships to standard string links.

Definition 6.3.3. Two $\mathbb{Z}$-tangles ℓ and ℓ' are *finitely isotopic* if, for every finite subset of $\mathbb{Z}$, the string links obtained from ℓ and ℓ' using only those strands are isotopic.

Any pair of weakly isotopic $\mathbb{Z}$-tangles is also finitely isotopic, since the movement of a fixed finite number of strings over the course of the weak isotopy can be realized for those strings by a strong isotopy. Verifying that the converse also holds would be an interesting question for future exploration.

6.4 Planar Knotoids

The term "knotoid" is sometimes used more generally than the way we use it here; what we call "knotoids" above might be called "knotoids on S^2". This is in reference to the fact that S^2 appears as the ambient space in the definition of knotoid diagrams, but alternative ambient surfaces can also be chosen. A common alternative choice is $\mathbb{R}^2$, giving rise to *knotoids on* $\mathbb{R}^2$, also called *planar knotoids* ([49], Section 10).

In the context of planar knotoids, there are equivalents to $\overline{A}$ and A, but we will see that there is no completely straightforward counterpart to Bh.

Fix a genus-2 handlebody $\overline{M}$, two distinct points w_0, w_1, on $\partial\overline{M}$, and an element ζ of the relative homology group $H_1(\overline{M}, \{\overline{w}_0, \overline{w}_1\})$ with $\partial\zeta = \overline{w}_1 - \overline{w}_0$. (Note that any two choices of $\overline{M}$, $\overline{w}_0$, $\overline{w}_1$, and ζ are essentially equivalent in that between the two choices of $\overline{M}$ there is a diffeomorphism that respects the other choices.) Then let $\overline{A_2}$ be the set of ∂-proper embeddings $I \to \overline{M}$ for which -1 and 1 respectively map to $\overline{w}_0$ and $\overline{w}_1$ and for which the embedding represents ζ. By [27] and [28] (Proposition 1), isotopy in $\overline{A_2}$ coincides with compositions with boundary-fixing automorphisms of M.

Given a planar knotoid, we can produce an element of $\overline{A_2}$: After picking a diagram, let us perform an ambient isotopy of $\mathbb{R}^2$ to put the endpoints into a standard position (say, $(0,-1)$ and $(0,1)$), then do another ambient isotopy just in a neighborhood of the endpoints to normalize winding numbers, then next delete small open disks around the two endpoints and the point at infinity to get a closed genus-0 surface with three boundary components, and then finally take the product of that surface with I to define $\overline{M}$. The embedding of I is determined by lifting the knotoid diagram into the thickened surface using the crossing information. The endpoints $\overline{w}_0$ and $\overline{w}_1$ as well as the class ζ can be inferred from the embedding. None of our choices affect the isotopy class of the resulting element of $\overline{A_2}$.

Note that the above process doesn't generalize to knotoids on punctured spheres with more than one puncture, because the two endpoints give us only two degrees of freedom in the first homology and also because of the potential for braiding of the endpoints.

Of course, the fundamental group of $\overline{M}$ is the free group $\mathbb{Z} * \mathbb{Z}$. The universal cover M of $\overline{M}$ is isomorphic to a copy of a 3-ball minus a Cantor set on the boundary. To define $(\mathbb{Z} * \mathbb{Z})$-tangles, we will need to pick a fixed (based) homotopy class of representatives of ζ and a lift $w_0 \in \partial M$ of $\overline{w}_0$. Having made these choices, let us define a lift w_1 by picking a representative of the aforementioned homotopy class, lifting it to M in such a way that it starts at w_0, and letting w_1 be the other endpoint. The other lifts of $\overline{w}_0$ and $\overline{w}_1$ can be obtained by applying Deck transformations of M to w_0 and w_1. Let A_2 be the set of biproper embeddings $(\mathbb{Z}*\mathbb{Z}) \times I \to M$ satisfying $(u,-1) \mapsto u(w_0)$ and $(u,1) \mapsto u(w_1)$ for all $u \in \mathbb{Z}*\mathbb{Z}$. Then there is a Birman–Hilden map from the subset of isotopy classes in $\overline{A_2}$ for which the strand represents the chosen homotopy class to the set of isotopy classes in A_2.

A planar knotoid can be assigned a triple of unsigned heights, for example as in [2] (page 6). Each of these three unsigned heights has corresponding signed variants, and each satisfies its own version of Theorem 3.3.1. The proof of Theorem 3.3.1 carries over to this new context with little modification.

Bibliography

[1] Aaron Abrams. "Braids". In: *Office Hours with a Geometric Group Theorist*. Ed. by Matt Clay and Dan Margalit. Princeton, NJ: Princeton University Press, 2017.

[2] Colin Adams et al. "Generalizations of Knotoids and Spatial Graphs". 2022. URL: `https://arxiv.org/abs/2209.01922`.

[3] J.W. Alexander. "Topological Invariants of Knots and Links". In: *Transactions of the American Mathematical Society* 30.2 (1928), pp. 275–306.

[4] Javier Aramayona and Nicholas G. Vlamis. "Big Mapping Class Groups: An Overview". In: *In the Tradition of Thurston: Geometry and Topology*. Ed. by Ken'ichi Ohshika and Athanase Papadopoulos. New York: Springer Nature, 2020.

[5] Emil Artin. "Theory of Braids". In: *Annals of Mathematics* 48.1 (1947), pp. 101–126.

[6] Agnese Barbensi and Dimos Goundaroulis. "f-distance of Knotoids and Protein Structure". In: *Proceedings of the Royal Society A* 477.2246 (2021).

[7] Agnese Barbensi et al. "Double Branched Covers of Knotoids". In: *Communications in Analysis and Geometry* 30.5 (2022), pp. 1007–1057.

[8] Etienne Batelier and Julien Ducoulombier. "Operadic Actions on Long Knots and 2-String Links". In: *Algebraic & Geometric Topology* 23.2 (2023), pp. 833–882.

[9] Joan Birman and Hugh Hilden. "Lifting and Projecting Homeomorphisms". In: *Archiv der Mathematik* 23 (1972), pp. 428–434.

[10] Joan Birman and Hugh Hilden. "On Isotopies of Homeomorphisms of Riemann Surfaces". In: *Annals of Mathematics* 97 (1973), pp. 424–439.

[11] Joan Birman and Hugh Hilden. "On the Mapping Class Groups of Closed Surfaces as Covering Spaces". In: *Annals of Mathematics Studies* 66 (1971), pp. 81–115.

[12] Ryan Blair, John Burke, and Robin Koytcheff. "A Prime Decomposition Theorem for the 2-String Link Monoid". In: *Journal of Knot Theory and Its Ramifications* 24.2 (2015), p. 1550005.

[13] John Burke and Robin Koytcheff. "A Colored Operad for String Link Infection". In: *Algebraic & Geometric Topology* 15 (2015), pp. 3371–3408.

[14] James W. Cannon and Michael Starbird. "A Diagram Oriented Proof of Dehn's Lemma". In: *Topology and its Applications* 26 (1987), pp. 193–205.

[15] John H. Conway and Cameron Gordon. "Knots and Links in Spatial Graphs". In: *Journal of Graph Theory* 7.4 (1983), pp. 445–453.

[16] Ioannis Diamantis, Sofia Lambropoulou, and Sonia Mahmoudi. "Equivalence of Doubly Periodic Tangles". 2023. URL: `https://arxiv.org/abs/2310.00822`.

[17] Ryan Dickmann. "Mapping Class Groups of Surfaces with Noncompact Boundary Components". In: *Algebraic & Geometric Topology* 23.6 (2023), pp. 2777–2821.

[18] Paul Fabel. "The Mapping Class Group of a Disk with Infinitely Many Holes". In: *Journal of Knot Theory and Its Ramifications* 15.1 (2006), pp. 21–29.

[19] Benson Farb and Dan Margalit. *A Primer on Mapping Class Groups*. Princeton Mathematical Series 49. Princeton, NJ: Princeton University Press, 2011.

[20] Benedict Freedman and Michael H. Freedman. "Kneser-Haken Finiteness for Bounded 3-Manifolds, Locally Free Groups, and Cyclic Covers". In: *Topology* 37.1 (1998), pp. 133–147.

[21] Dimos Goundaroulis et al. "Topological Models for Open-Knotted Protein Chains Using the Concepts of Knotoids and Bonded Knotoids". In: *Polymers* 9.9 (2017).

[22] Mikhail Goussarov, Michael Polyak, and Oleg Viro. "Finite-Type Invariants of Classical and Virtual Knots". In: *Topology* 39 (2000), pp. 1045–1068.

[23] Neslihan Gügümcü and Louis Kauffman. "New Invariants of Knotoids". In: *European Journal of Combinatorics* 65 (2017), pp. 186–229.

[24] Neslihan Gügümcü and Louis Kauffman. "Parity, Virtual Closure and Minimality of Knotoids". In: *Journal of Knot Theory and Its Ramifications* 30.11 (2021), p. 2150076.

[25] Neslihan Gügümcü and Louis Kauffman. "Quantum Invariants of Knotoids". In: *Communications in Mathematical Physics* 387 (2021), pp. 1681–1728.

[26] Nathan Habegger and Xiao-Song Lin. "The Classification of Links up to Link-Homotopy". In: *Journal of the American Mathematical Society* 3.2 (1990), pp. 389–419.

[27] Allen Hatcher. "A Proof of the Smale Conjecture, $\mathrm{Diff}(S^3) \cong O(4)$". In: *Annals of Mathematics* 17 (1983), pp. 553–607.

[28] Allen Hatcher. "Homeomorphisms of Sufficiently Large P^2-Irreducible 3-Manifolds". In: *Topology* 15 (1976), pp. 343–347.

[29] Louis H. Kauffman. "Virtual Knot Theory". In: *Encyclopedia of Knot Theory*. Ed. by Colin Adams et al. Boca Raton, FL: CRC Press, 2021.

[30] Sera Kim, Young Ho Im, and Sunho Lee. "A Family of Polynomial Invariants for Knotoids". In: *Journal of Knot Theory and Its Ramifications* 27.11 (2018).

[31] Shin'ichi Kinoshita. "On Elementary Ideals of Polyhedra in the 3-Sphere". In: *Pacific Journal of Mathematics* 42.1 (1972), pp. 89–98.

[32] P.G. Korablev, Y.K. May, and V.V. Tarkaev. "Classification of Low Complexity Knotoids". In: *Siberian Electronic Mathematical Reports* 15 (2018), pp. 1237–1244.

[33] Philipp Korablev and Vladimir Tarkaev. "A Relation Between the Crossing Number and the Height of a Knotoid". 2020. URL: https://arxiv.org/abs/2009.02718.

[34] Deniz Kutluay. "Winding Homology of Knotoids". 2020. URL: https://arxiv.org/abs/2002.07871.

[35] W.B. Raymond Lickorish. *An Introduction to Knot Theory*. Graduate Texts in Mathematics 175. New York: Springer Science & Business Media, 1997.

[36] Larsen Linov. "Signed Heights of Knotoids". In: *Journal of Knot Theory and Its Ramifications* 31.5 (2022), p. 2250037.

[37] Charles Newton Little. "XXX.—Non-Alternate ± Knots". In: *Transactions of the Royal Society of Edinburgh* 39 (1900), pp. 771–778.

[38] Wilhelm Magnus. "Beziehungen Zwischen Gruppen und Idealen in Einem Speziellen Ring". In: *Mathematische Annalen* 111 (1935).

[39] Dan Margalit and Rebecca Winarski. "Braid Groups and Mapping Class Groups: The Birman–Hilden Theory". In: *Bulletin of the London Mathematical Society* 53 (2021), pp. 643–659.

[40] Darryl McCullough. "Homeomorphisms Which Are Dehn Twists on the Boundary". In: *Algebraic & Geometric Topology* 6 (2006), pp. 1331–1340.

[41] John Milnor. "Link Groups". In: *Annals of Mathematics* 59 (1954), pp. 177–195.

[42] Wout Moltmaker. "Framed Knotoids and Their Quantum Invariants". In: *Communications in Mathematical Physics* 393 (2022), pp. 1035–1061.

[43] Christos D. Papakyriakopoulos. "On Dehn's Lemma and the Asphericity of Knots". In: *Annals of Mathematics* 66.1 (1957), pp. 1–26.

[44] John Pardon. "The Hilbert–Smith Conjecture for Three-Manifolds". In: *Journal of the American Mathematical Society* 26.3 (2013), pp. 879–899.

[45] Derek Robinson. *A Course in the Theory of Groups*. Second Edition. Graduate Texts in Mathematics 80. New York: Springer Science & Business Media, 1996.

[46] John Stallings. "Homology and Central Series of Groups". In: *Journal of Algebra* 2 (1965), pp. 170–181.

[47] Ryan Stees. "Milnor's Invariants for Knots and Links in Closed Orientable 3-Manifolds". PhD thesis. Indiana University, 2023.

[48] Vladimir Tarkaev. "A Prime Decomposition Theorem for String Links in a Thickened Surface". 2024. URL: https://arxiv.org/abs/2403.12492.

[49] Vladimir Turaev. "Knotoids". In: *Osaka Journal of Mathematics* 49 (2012), pp. 195–223.

[50] Friedhelm Waldhausen. "On Irreducible 3-Manifolds Which are Sufficiently Large". In: *Annals of Mathematics* 87.1 (1968), pp. 56–88.

[51] Eric W. Weisstein. *Tait's Knot Conjectures*. From *MathWorld*. 2024. URL: https://mathworld.wolfram.com/TaitsKnotConjectures.html.

[52] Yi Xie. "Instantons and Annular Khovanov Homology". In: *Advances in Mathematics* 388 (2021), p. 107864.